

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 2003, 09:43:59 ; Search time 42 seconds
(without alignments)
826.593 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1886

Sequence: 1 MNKILFFLFFSSVLETFFA.....EAGQYMGNIWTFPSSQTL 361

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1558.5	82.6	360	D56617	cfaf protein precu
2	949.5	50.3	363	S49539	Cold protein precu
3	930.5	49.3	364	S57937	Cold protein precu
4	275.5	14.6	359	AE0541	probable fimbrial
5	120	6.4	1335	G90975	probable factor (1
6	120	6.4	2660	S28822	probable invasin 2
7	114	6.0	650	S22835	alpha-aggutinin -
8	109	5.8	2383	S55097	probable membrane
9	108	5.7	524	S55097	probable membrane
10	108	5.7	2529	B64635	toxin-like outer m
11	107	5.7	304	A64904	probable fimbrial
12	107	5.7	304	C90892	probable adhesin (
13	107	5.7	304	S85725	probable adhesin (
14	106	5.6	671	A38109	autolysin - Entero
15	105.5	5.6	843	AC2507	hypothetical prote
16	105	5.6	598	AB1236	internalin protein
17	104.5	5.5	1441	B86807	hypothetical prote
18	103.5	5.5	691	B75622	hypothetical prote
19	103.5	5.5	1269	A90267	proteinase related
20	103.5	5.5	2399	H71879	toxin-like outer m
21	102	5.4	692	G90284	hypothetical prote
22	101	5.4	1238	AH0038	probable exported
23	100	5.3	1752	C83339	hypothetical prote
24	99.5	5.2	856	AB1580	phenylalanyl-tRNA
25	99	5.2	802	B81399	probable periplasm
26	99	5.2	1849	C41859	Iga-specific metal
27	98	5.2	608	H64473	hypothetical prote
28	98.5	5.2	4919	T31105	hypothetical prote

30	98	5.2	587	2	AC1510	internalin protein
31	98	5.2	758	2	S47866	SP21 protein - ye
32	98	5.2	1014	2	S37405	cytotoxic necrotiz
33	98	5.2	1157	2	S38160	NUP133 protein - y
34	97.5	5.2	1797	2	F69195	cell surface glyco
35	97	5.1	454	2	T01337	hypothetical prote
36	96.5	5.1	2364	2	I40884	cytotoxin L - Clos
37	96	5.1	992	2	T28421	probable DNA-direc
38	95.5	5.1	713	1	AB85XR	cyclomaldextrin
39	95.5	5.1	895	2	T02597	Mutator-like trans
40	95	5.0	413	1	S28602	translation releas
41	95	5.0	561	2	AB4113	transposase (12) B
42	95	5.0	2334	2	S32920	cell wall-associat
43	95	5.0	5188	2	B85547	probable RTX fam1
44	94.5	5.0	835	2	E71691	outer membrane ass
45	94.5	5.0	858	2	T47223	replication licens

ALIGNMENTS

RESULT 1

D56617 cfaf protein precursor - Escherichia coli plasmid NTP113

C:Species: Escherichia coli

C:Date: 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 10-Dec-1999

C:Accession: D56617 R:Jordl, B.J.; Willshaw, G.A.; van der Zelfst, B.A.; Gaasstra, W.

DNA Seq. 2, 257-263, 1992

A:Title: The complete nucleotide sequence of region 1 of the CFA/I fimbrial operon of

A:Reference number: A56617; MUID:92329981; PMID:1352712

A:Accession: D56617

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-360 <OR>

A:Cross-references: GB:M55661; NID:9145507; PIDN:AA41417.1; PID:9145511

A:Experimental source: enterotoxigenic strain, CFA/I-ST plasmid NTP113

A>Note: sequence extracted from NCBI backbone (NCBIN:108960, NCBI:P108972)

C:Genetics:

A:Gene: cfaf

A:Genome: plasmid

C:Superfamily: Escherichia colonizing factor antigen cfaf

Query Match 82.6%; Score 1558.5; DB 2: Length 360;

Best Local Similarity 81.2%; Pred. No. 1.4e-109;

Matches 293; Conservative 31; Mismatches 36; Indels 1; Gaps 1;

QY	1	MNKLFFLFFSSVLETFFAVSADKIPGDEITNIFGPRDRNSSPKHNLNNHITAYSE	60
DB	1	MNKLFFLFFSSVLETFFAVSADKIPGDEITNIFGPRDRNSSPKHNLNNHITAYSE	60
QY	61	SHLDRMTFLCLSSNITNGACPISENSSSSSEETITITQTEKRRLIRELOIKY	120
DB	61	SHLDRMTFLCLSSNITNGACPISENSSSSSEETITITQTEKRRLIRELOIKY	120
QY	61	SHLDRMTFLCLSSNITNGACPISENSSSSSEETITITQTEKRRLIRELOIKY	120
DB	61	SHLDRMTFLCLSSNITNGACPISENSSSSSEETITITQTEKRRLIRELOIKY	120
QY	121	KOLFEKSVCPGSLTNSAHFPCNKMAAGASLYITIPGELKNLPFGGIMATTLARK	180
DB	121	KOLFEKSVCPGSLTNSAHFPCNKMAAGASLYITIPGELKNLPFGGIMATTLARK	180
QY	121	KOLFEKSVCPGSLTNSAHFPCNKMAAGASLYITIPGELKNLPFGGIMATTLARK	180
DB	121	KOLFEKSVCPGSLTNSAHFPCNKMAAGASLYITIPGELKNLPFGGIMATTLARK	180
QY	181	RRYSSETYGTITNITIKLTDKGNIOIWLPOFKSDARVDLNLPTGGGTYIGRNSVDMCPY	240
DB	181	RRYSSETYGTITNITIKLTDKGNIOIWLPOFKSDARVDLNLPTGGGTYIGRNSVDMCPY	240
QY	181	RRYSSETYGTITNITIKLTDKGNIOIWLPOFKSDARVDLNLPTGGGTYIGRNSVDMCPY	240
DB	181	RRYSSETYGTITNITIKLTDKGNIOIWLPOFKSDARVDLNLPTGGGTYIGRNSVDMCPY	240
QY	241	DGYSTNSSLETRFDONNKSQKRYLKIINDTKEIATLSLLAGKSLFTPTNGSLNI	300
DB	241	DGYSTNSSLETRFDONNKSQKRYLKIINDTKEIATLSLLAGKSLFTPTNGSLNI	300
QY	241	DGYSTNSSLETRFDONNKSQKRYLKIINDTKEIATLSLLAGKSLFTPTNGSLNI	300
DB	241	DGYSTNSSLETRFDONNKSQKRYLKIINDTKEIATLSLLAGKSLFTPTNGSLNI	300
QY	301	ADAASLETWNRITVATVMEISVPYLCWGRQLDAKVENPAGQYMGNIWTFPSSQT	360
DB	301	ADAASLETWNRITVATVMEISVPYLCWGRQLDAKVENPAGQYMGNIWTFPSSQT	360
QY	301	ADAASLETWNRITVATVMEISVPYLCWGRQLDAKVENPAGQYMGNIWTFPSSQT	360
DB	301	ADAASLETWNRITVATVMEISVPYLCWGRQLDAKVENPAGQYMGNIWTFPSSQT	360
QY	361	L 361	

DB 360 L 360

RESULT 2

S49539

Cood protein precursor - Escherichia coli

C:Species: Escherichia coli

C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 10-Dec-1999

C:Accession: S49539

R:Frederich, B.J.; Karakashian, A.; Melsen, L.R.; Wakefield, J.C.; Scott, J.R.

Mol. Microbiol. 12, 387-401, 1994

A:Title: Cood and Cood are required for assembly of CS1 pill.

A:Reference number: S49538; MUID:94344028; PMID:7915003

A:Accession: S49539

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-363 <FRO>

A:Cross-references: EMBL:X76908; NID:9488735; PIDN:CA454230.1; PID:9488737

Superfamily: Escherichia colonizing factor antigen cfae

Query Match 50.3%; Score 949.5; DB 2; Length 363;

Best Local Similarity 53.1%; Pred. No. 7.7e-64;

Matches 197; Conservative 55; Mismatches 98; Indels 21; Gaps 11;

```

OY 3 KILFETLFSSVLETFVAVSADKIPGDE--STTNIF-GPR-DRNSSPKHNTLNHITAY 58
DB 2 KILFETLFSSVLETFVAVSADKIPGDE--STTNIF-GPR-DRNSSPKHNTLNHITAY 57
OY 59 SESHTLYDRMTFLCSHNTLNGACPTSENPSSSSVS-GETNITQTEKRSILKRELOI 117
DB 58 SLSHSLYDRIVFLCTSSSNPNVNGACPTI---CTSGVOGTITITITQTEKRSILKRNINL 114
OY 118 KGYKOLLKSVNCPGSG--LTLSAHFNCNKA-ASGASLYIYIPAGELKNLPGGIMDAT 174
DB 115 AGNKKPIWENSCDFSNLMVLNLSKSGAHGANGTLLNXYIIPAGEINKLPGGIMWAT 174
OY 175 LKLRKRRYSST---VGTPTNTITIKLTDKGNIOIWLPOKSDARVULNLRPTGGIYI 230
DB 175 LKLRKRRYSST---VGTPTNTITIKLTDKGNIOIWLPOKSDARVULNLRPTGGIYI 230
OY 231 GRNSVDMCFYDGYSTNSSLEIRFQDNNPKSPDKFYLRKINDTEKEIAYTSLLAGKSL 290
DB 234 GSNLDMCFYDGYSTNSSLEIRFQDNNPKSPDKFYLRKINDTEKEIAYTSLLAGKSL 292
OY 291 TPTNCTSLNADASLEFNMMNRITAVTPEISVPLVCPGRQLQDAKVENPEAGGYMGI 350
DB 293 YPVNQSFTINDSVLETFNMNRITAVTPEISVPLVCPGRQLQDAKVENPEAGGYMGI 352
OY 351 NTTPSSQTL 361
DB 353 YTFPVSVENTL 363

```

RESULT 3

S57937

Cod protein precursor - Escherichia coli

C:Species: Escherichia coli

C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 10-Dec-1999

C:Accession: S57937

R:Frederich, B.J.; Karakashian, A.; Melsen, L.R.; Scott, J.R.

submitted to the EMBL Data Library, January 1995

A:Description: The genes for CS2 pill of enterotoxigenic Escherichia coli and their inte

A:Reference number: S57934

A:Accession: S57937

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-364 <FRO>

A:Cross-references: EMBL:Z47800; NID:9897725; PIDN:CA48763.1; PID:9897729

C:Genetics:

A:Start codon: TTG

C:Superfamily: Escherichia colonizing factor antigen cfae

Query Match 49.3%; Score 930.5; DB 2; Length 364;

Best Local Similarity 48.4%; Pred. No. 2.1e-62;

Matches 177; Conservative 64; Mismatches 118; Indels 7; Gaps 3;

```

OY 1 MKKILFETLFSSVLETFVAVSADKIPGDE--STTNIF-GPR-DRNSSPKHNTLNHITAY 58
DB 1 MKKILFETLFSSVLETFVAVSADKIPGDE--STTNIF-GPR-DRNSSPKHNTLNHITAY 60
OY 59 SESHTLYDRMTFLCSHNTLNGACPTSENPSSSSVS-GETNITQTEKRSILKRELOI 118
DB 61 NKDHSLEDRMTFLCSHNTLNGACPTSENPSSSSVS-GETNITQTEKRSILKRELOI 118
OY 119 GYKOLLKSVNCPGSG--LTLSAHFNCNKA-ASGASLYIYIPAGELKNLPGGIMDAT 175
DB 119 GYKOLLKSVNCPGSG--LTLSAHFNCNKA-ASGASLYIYIPAGELKNLPGGIMDAT 178
OY 176 KLRVRRYSSTGYTNTITIKLTDKGNIOIWLPOKSDARVULNLRPTGGIYI 235
DB 179 ELRVRRHNDYNGYKAVITVDLTPKGNIOIWLPOKSDARVULNLRPTGGIYI 238
OY 236 DMCFTYDGYSTNSSLEIRFQDNNPKSPDKFYLRKINDTEKEIAYTSLLAGKSLPTNG 295
DB 239 EMCLTDGYSTHSQSIEMRFQDNNPKSPDKFYLRKINDTEKEIAYTSLLAGKSLPTNG 298
OY 296 TSLNADASLEFNMMNRITAVTPEISVPLVCPGRQLQDAKVENPEAGGYMGI 355
DB 299 EAPFTINDSVLETFNMNRITAVTPEISVPLVCPGRQLQDAKVENPEAGGYMGI 358
OY 356 PSSQTL 361
DB 359 PSSSL 364

```

RESULT 4

AE0541

probable fimbrial protein tcfD [imported] - Salmonella enterica subsp. enterica serov

C:Species: Salmonella enterica subsp. enterica serovar typhi

A:Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AE0541

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church

th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr

, S.; Moulé, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se

A:Reference number: AE0502; MUID:21534947; PMID:11677608

A:Accession: AE0541

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-359 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08773.1; PID:916501589; GSPDB:GN00176

C:Genetics:

A:Gene: tcfD

Query Match 14.6%; Score 275.5; DB 2; Length 359;

Best Local Similarity 29.7%; Pred. No. 3.2e-13;

Matches 90; Conservative 48; Mismatches 130; Indels 35; Gaps 11;

```

OY 70 FLCLSHNTLNGACPTSENPSSSVSGETNITQTEKRSILKRELOIKYKOLLKFSVN 129
DB 79 WCNRSNNRNBACETHLWVWYAFGAVSKRLRREDSIAEITL-----ILGSVR 131
OY 130 --CEGTLNSAHFNCNKAAS--GASLYIYIPAGELKNLPGGIMDATLKL-RVKRRY 183
DB 132 DACTGTG-----INNNAACQWGRSLKRIPESEELAKIPTSGTWKATLVLDLQGG 183
OY 184 SETGYTNTITIKLTD--KGNIOIWLPOF-KSDARVDLNRPTGGIYIGRNSVDGCFY 240
DB 184 DDLPGTSTDTLWVTDHFAENNAIYFPGCTATPRVDLNLHRNNAASQSGRANIDCLY 243
OY 241 DGYSTNSGLEIRQDNNPKSGDKFYLRKINDTEKEIAYTSLLAGKSLPTNGTSLNI 300
DB 244 DG-GVKARSLQMKIEGSKSGTGPVYIKSDADT--IDYAVSMYNGRSIPYTRGVESFL 300

```

QY 301 ADAASLETNNRITAVTMEISVPLVLCMPGRIOLDAK---VENPEAGOVYMNINVTTPS 357
 Db 301 DNVDKRAATR-----PVVLEQROAVRCVPLVLTTPFPENIREKRSSEYGTGLVTWLMG 355
 QY 358 SQP 360
 Db 356 TOT 358

RESULT 5

G90975

Probable factor [imported] - Escherichia coli (strain O157:H7, substrain RMD 0509952)

C/Species: Escherichia coli

C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001

C/Accession: G90975

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001

A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic reference number: A99629; MUID:21156231; PMID:11258796

Accession: G90975

Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-1335 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA036198.1; PID:g13362243; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RMD 0509952

C:Genetics:

A:Gene: ECS2775

Query Match 6.4%; Score 120; DB 2; Length 1335;
 Best Local Similarity 20.4%; Pred. No. 0.86;
 Matches 80; Conservative 66; Mismatches 156; Indels 90; Gaps 15;

QY 13 SSVLEFFNVSADKIPGDESTINIFGPRDRNESSPKHNILNNHTAY---SESHLIDRMT 69
 Db 746 SAKIATLSASNNCVLANENNAANTVSVADEGS---NPINDHTVTFVAVLSGATSFNON 802
 QY 70 FLCLSHNTLNGACPTSENPSSSVSGETNITLOTFEKRSLI-----K 112
 Db 803 ---TAKTDVNGLA-TFDLKSSKOEDNTVETLENGVKQTLIVSPGDSSTAQVDLOKSK 857
 QY 113 RELQIKGY-KQLLEKSYNCPGSLTNSAHFNCKRNAASGASLYLIPAGELKNLPPGWI 171
 Db 858 NEVVADGNDGATGATVADKGNLNDVKTFVNSAAKLSQTEVNSHD-----GIA 910
 QY 172 DATLKLVRKRYSETGTITNTIKLTDKGNQIQLPQKRSARVDLNRPGGGYIG 231
 Db 911 TATLT-----SLKNGDYVTASVSSGQANQOVIIFIGDSTAAULTLSV-PSGDITV-- 960
 QY 232 RNSVDMCFYDGYSTNSSLE---IRFODNPKSDGKFLYLRKINDDTKEIAYTLISLLAG 287
 Db 961 -----TNTAPLHMTATLDKKNPDKKEITFSVND-----VASRFSISNSG 1003
 QY 288 KSLTPINGSL-----NIADAASLETNNRITAVTMEISVPLC 327
 Db 1004 KGMTDSNGTAIASLTGLAGTHMITARLANSVSDQPMTFVADKRAVVAVLQTSKAEII 1063
 QY 328 WPG--RLQDRAKVENPEAGOVYMNINVTTPS 357
 Db 1064 GNGVDETTLTATYKDP-FDNVKNLSVVERTS 1094

RESULT 6

E85822

Probable Invasin Z3135 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)

C/Species: Escherichia coli

C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C/Accession: E85822

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew Miller, L.; Gotohck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoussis, K.; Apodaca, Nature 409, 529-533, 2001

A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: E85822
 A>Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-2660 <STO>
 A:Cross-references: GB:AE005174; NID:g12516151; PIDN:AA057041.1; GSPDB:GN00145; UWGP:
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: Z3135

Query Match 6.4%; Score 120; DB 2; Length 2660;
 Best Local Similarity 20.4%; Pred. No. 2.1;
 Matches 80; Conservative 66; Mismatches 156; Indels 90; Gaps 15;

QY 13 SSVLEFFNVSADKIPGDESTINIFGPRDRNESSPKHNILNNHTAY---SESHLIDRMT 69
 Db 737 SAKIATLSASNNCVLANENNAANTVSVADEGS---NPINDHTVTFVAVLSGATSFNON 793
 QY 70 FLCLSHNTLNGACPTSENPSSSVSGETNITLOTFEKRSLI-----K 112
 Db 794 ---TAKTDVNGLA-TFDLKSSKOEDNTVETLENGVKQTLIVSPGDSSTAQVDLOKSK 848
 QY 113 RELQIKGY-KQLLEKSYNCPGSLTNSAHFNCKRNAASGASLYLIPAGELKNLPPGWI 171
 Db 849 NEVVADGNDGATGATVADKGNLNDVKTFVNSAAKLSQTEVNSHD-----GIA 901
 QY 172 DATLKLVRKRYSETGTITNTIKLTDKGNQIQLPQKRSARVDLNRPGGGYIG 231
 Db 902 TATLT-----SLKNGDYVTASVSSGQANQOVIIFIGDSTAAULTLSV-PSGDITV-- 951
 QY 232 RNSVDMCFYDGYSTNSSLE---IRFODNPKSDGKFLYLRKINDDTKEIAYTLISLLAG 287
 Db 952 -----TNTAPLHMTATLDKKNPDKKEITFSVND-----VASRFSISNSG 994
 QY 288 KSLTPINGSL-----NIADAASLETNNRITAVTMEISVPLC 327
 Db 995 KGMTDSNGTAIASLTGLAGTHMITARLANSVSDQPMTFVADKRAVVAVLQTSKAEII 1054
 QY 328 WPG--RLQDRAKVENPEAGOVYMNINVTTPS 357
 Db 1055 GNGVDETTLTATYKDP-FDNVKNLSVVERTS 1085

RESULT 7

S22835

alpha-agglutinin - yeast (Saccharomyces cerevisiae)

N/Alternate names: 22k glycoprotein; protein J1418; protein YJR004c

C/Species: Saccharomyces cerevisiae

C/Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 29-Oct-1999

C/Accession: S22835; S51229; A32822; S55192; S57019

R:Husser, K.; Tanner, W.

FEBS Lett. 255, 290-294, 1989

A/Title: Purification of the inducible alpha-agglutinin of S. cerevisiae and molecule

A:Reference number: S22835; MUID:90005993; PMID:2676603

A:Accession: S22835

A:Molecule type: DNA

A:Residues: 1-650 <HAY>

A:Cross-references: EMBL:X16861; NID:g3352; PIDN:CAA34752.1; PID:g3353

A:Accession: S51229

A:Molecule type: protein

A:Residues: 20-24 <HA2>

R:Lipke, P.N.; Wojciechowski, D.; Kurjan, J.

Mol. Cell. Biol. 9, 3155-3165, 1989

A/Title: AG-alpha-1 is the structural gene for the Saccharomyces cerevisiae alpha-299

A:Reference number: A32822; MUID:90014768; PMID:2677666

A:Accession: A32822

A:Molecule type: DNA

A:Residues: 1-448, 'P', 450-555, 'E', 557-580, 'L', 582-650 <LIP>

A:Cross-references: GB:M8164; NID:g171041; PIDN:AAA34417.1; PID:g171044

R:de Haan, M.; Smits, P.H.M.; Grivell, L.A.

submitted to the EMBL Data Library, May 1995

A:Reference number: S55183

A:Accession: S55192

A.Molecule type: DNA
 A.Residues: 1-650 <DEH>
 A.Cross-references: EMBL:X87611; NID:g854567; PIDN:CAA60926.1; PID:g854577
 Ride Haan, M.; Griwell, L.A.; Smits, P.H.M.
 submitted to the Protein Sequence Database, September 1995
 A.Reference number: S56771
 A.Accession: S57019
 A.Molecule type: DNA
 A.Residues: 1-650 <ZAG>
 A.Cross-references: EMBL:249504; NID:g1015625; PIDN:CAA89526.1; PID:g1015626; MIPS:YUR00
 C.Genetics:
 A.Gene: SGD:SAG1; AGAL1; AGALPHA1
 A.Cross-references: SGD:S0003764; MIPS:YUR004c
 A.Map position: 10R
 C.Keywords: glycoprotein

Query Match 6.0%; Score 114; DB 2; Length 650;
 Best Local Similarity 20.9%; Pred. No. 0.95; Mismatches 147; Indels 172; Gaps 23;
 Matches 98; Conservative 52;

4 ILFIPLFFSVL-----FTFA-----VSADKIPG-----DESITNIFGRDRNE--- 43
 8 ILMLRSLALASAININDITFSNLEITPLANKQPDGWTATFDFSLADASSIREDEFTL 67
 44 SSP---KNNILNNHITA-----YSESHLYDRMTFLC-----LSSHNTLN 80
 68 SMPHYRIKILNNSQTATISLADGTEAKCYVSOQAAYLYENTFTCTAQNDSLSSYNTD 127
 81 GACPTSENPSSSVSGENITLQTEKRSLLIREQINGKQLLRKSYNC-PSGLTLNSA 139
 128 GSITSLNFSDDGSSYEEL-----ENAKFKSGMLVKLGOMDVNFDPAATEENF 182
 140 HENCKNMAASGASLYL-YIPAGELKNLPFGGIMDATLKRVRSEYETGYTINIRIK 197
 183 HSGRSTGYGSPFSEYHLYGKCPNGYF---LGS---TEKI-----DYSSNNNVLD- 225
 198 LTKDNINQI-----WLPQKSDARVDLNLRTGGGTYIGRN---SVMCKFYDG--- 242
 226 --DCSSVQVYSSNDENDMWFQSYNDTNADV-----TCGSMWMTLDEKLYDGEML 275
 243 -----YSTNSSSLE--TRFODNN--PKSDGKF 265
 276 WYNALQSLPANVNTIDHLEFOYTOLODTIANTTYATQSTIREFYVYGRNLGTSASAKS 335
 266 YLRKINDTKET---AYTL-----LLAGKSLRPTNGSLNIADAA 304
 336 FISTTTDLTSLINTSAYSTGSIYETGNRTSEVISHVVTSTKLSPATSLTIAQTS 395
 305 SLETFWN-----RITAVMPETISVPYLCPPGRL 332
 396 IYSDSNITVGTDIHTTSEVSDVETISRETASTVVAAPTSTGTGTGAM 444
 DB

RESULT 8

D64962
 probable membrane protein b1978 - Escherichia coli (strain K-12)
 C.Species: Escherichia coli
 C.Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
 C.Accession: D64962
 R.Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A.Title: The complete genome sequence of Escherichia coli K-12.
 A.Reference number: A64720; MUID:97426617; PMID:9278503
 A.Accession: D64962
 A.Molecule type: DNA
 A.Status: nucleic acid sequence not shown; translation not shown
 A.Molecule type: DNA
 A.Residues: 1-2383 <BLAT>
 A.Cross-references: GB:AE000289; GB:U00096; NID:g1788285; PIDN:AACT5042.1; PID:g1788288;
 A.Experimental source: strain K-12, substrain MG1655
 C.Keywords: nucleotide binding; P-loop; transmembrane protein
 F:54-70/Domain: transmembrane #status predicted <TM>
 F:1564-1571/Region: nucleotide-binding motif A (P-loop)

Query Match 5.8%; Score 109; DB 2; Length 2383;
 Best Local Similarity 20.0%; Pred. No. 12;
 Matches 78; Conservative 65; Mismatches 157; Indels 90; Gaps 15;

13 SSVLETFPAVSADKIPGDESTINIFGRDRNESSPKHNLNNHITAY---SESHLYDRMT 69
 762 SAKIATLSASNNGVLANENAAVTASVNADEGS---NPINDHTVFAVLGSAFNSFNQN 818
 70 FLCLSHMTLNCACPTSENPSSSVSGENITLQTEKRSLL-----K 112
 819 ----TAKIDVNGLA-TFDLAKSKQEDNVEYVTELENGVQTLIVSGSSTRAQDLQSK 873
 113 RELQIKYKQL-LFKSVNCPGSLTNSAHFNCNKNAASGASLYLIPAGELKNLPFGGIW 171
 874 NEVVADGSDVTMTATVADAKGNLNDVMTFENVASAKLSQTEVNSHD-----GIA 926
 172 DATLKLVRKRRSEYETGYTINITYIKLTKDKNGIQIWLDPQKSDARVDLNLRTGGGTYIG 231
 927 TATLT-----SLKNGDYRVATASVSSGSOAQVNFIGDQSTALTLTSLV-PSGDITV-- 976
 232 RNSVDMCFYDGYSTNSS---SLERFODNPKSDGKFRYLKRIINDTEKIAVYTLISLAG 287
 977 -----TNTAQYMTATILQDKNGNPKDKETTFVYND-----VASKSISNGG 1019
 288 KSLPTNGTSL-----NIADAASLETNWNRTAVTPEISVPYLC 327
 1020 KGMTDSNGVALASLNGTLAGTHIMARLANSVSDAQMTFVADKRAVVVVLQTSKAEII 1079
 328 WFG--RLQIDAKVENPEAGQIMGNINVTFT 355
 1080 GNGVDETTLTATVKDP-SNHPPAGITVNT 1108
 DB

RESULT 9

S55097
 probable membrane protein YMR215w - yeast (Saccharomyces cerevisiae)
 N.Alternate names: hypothetical protein YMR261.09
 C.Species: Saccharomyces cerevisiae
 C.Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
 A.Accession: S55097
 R.Dedman, K.; Brown, D.; Bowman, S.
 submitted to the EMBL Data Library, June 1995
 A.Reference number: S55089
 A.Accession: S55089
 A.Molecule type: DNA
 A.Residues: 1-524 <DED>
 A.Cross-references: EMBL:249809; NID:g854459; PID:g854467; GSPDB:GN00013; MIPS:YMR215
 A.Experimental source: strain AB972
 C.Genetics:
 A.Gene: MIPS:YMR215w
 A.Cross-references: SGD:S0004828
 A.Map position: 13R
 C.Superfamily: glycopospholipid-anchored surface glycoprotein GAS1
 C.Keywords: transmembrane protein
 F:7-23/Domain: transmembrane #status predicted <TM>

Query Match 5.7%; Score 108; DB 2; Length 524;
 Best Local Similarity 24.7%; Pred. No. 2;
 Matches 69; Conservative 42; Mismatches 96; Indels 72; Gaps 18;

119 GYKQL-LFKSVNCPGSLTNSAHFNCNKNA--SGASLYLYIPAGELKNLPFGGIWDA 173
 264 GYDKLNTSFEDAVIP---LIFEYGCNKNPRTFDEVSEELY---GGLKNVFGSL--- 313
 174 TLKLVRKRRSEYETGYTINITYIKLTKDKNGIQIWLDPQKSD-ARVDNLNR---PTGGCT 228
 314 -----VEYTEEANNVGL---VKLDSGSL-----TKDPLVWLESQLKAVSLPTKES 359
 DB
 229 YIGRNSVDMC-----FYDGYSTNSSSLE-----IRFODNPKSDSK----- 264
 DB EISDSIYKCNNSAITNIYSGFTNNFLPQPAEIANMIYGVNGTNT-GKILLDVAVP 418


```

Oy 265 ----FLRKIRINDTKRIATLTLGLAGSLPNTGTSINDADSLNNMR-TEATVM 31.9
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 419 TTFNTYIKKNNKDF--ISATISYDRA-NSLNELDVATTAAYKASTSSSSRSLTSTSP 475

Oy 320 EISVPLCMPGRLOLDKAYENPEACQWGTI-NVTFPS 357
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 476 SSSSTGSSSTG----SSASSSSKSKGQVGNIVNVSFSQS 510

```

[illegible]

```

A:Reference number: A38109; MUID:91358349; PMID:1679432
A:Accession: A38109
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-671 <BEL>
A:Cross-references: GB:M58002; NID:915368; PID:AAA67325.1; PID:9829194

Query Match          5.6%; Score 106; DB 2; Length 671;
Best Local Similarity 23.6%; Pred. No. 3.9; Mismatches 103; Indels 94; Gaps 17;
Matches 73; Conservative 39;

Oy 41 RNESSPKHNT-LNNHITAYSESHLYDREMTFLCSSHNTLNGACPTSENPSSSV----- 94
Db 309 RYATDPSEYNAKLNNVITAY--NLFOYDPPSSGGMTGGGTGATPGTSGNNQSGTWTYYTVK 366
Oy 95 SGER--NTLTF-----TEKRSL--IKRRLQIKYKQLLFKSVNCPSGTLTNSAHNCKN 146
Db 367 SGGTLNKLAAOYGVSVANLRSMNGISGDLIFVGOKLIYKKGA---SGTGGSGNGSGNN 423
Oy 147 AASASLYLYPAGELKLPFGGIMDAFLKLRVRRSEYGTGTYINTITLTKGNIOI 206
Db 424 -QSGTNTFYTKSGDITL-----KIAOYG-----VTY-----ANLRS 455
Oy 207 WLPOPKSDARVDLNL-----RPTGGGYTIGRNSVDMCFYDGTSTNSSLEIRFODNN 258
Db 456 W-----NMSGDLIFVGOKLIYKKGTSGNT-----GGSSNGS-----NNN 491
Oy 259 PKSGKFEYLRKINDPKRIA--YTLSS-----TLGKSKSLPTNGTSLNTADA 303
Db 492 QSGTNTYTYTKSGTLNKIAAOYGVSVANLRSMNGISGDLIFAGOKIIVKKGTSGNTGGS 551
Oy 304 ASLETNMNR 312
Db 552 SNGSGNNNQ 560

RESULT 15
AC2507
hypothetical protein al17235 [imported] - Nostoc sp. (strain PCC 7120) plasmid PCC712121
C:Species: Nostoc sp. PCC 7120
A:Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kunitz, T.; Sasamoto, S.; Watanabe, A.; Iriq
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2507
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-843 <KUR>
A:Cross-references: GB:BA000020; PIDN:BA878319.1; PID:917135773; GSPDB:GN00180
A:Experimental source: Strain PCC 7120
A:Genetics:
A:Gene: al17235
A:Genome: plasmid

Query Match          5.6%; Score 105.5; DB 2; Length 843;
Best Local Similarity 18.9%; Pred. No. 5.8;
Matches 71; Conservative 67; Mismatches 127; Indels 111; Gaps 18;

Oy 77 NTLNGACPTSEN-----PSSSGVSGETNITLQ-----FT--EKRLIKREL.115
Db 297 NILDGSLVFTDNGHFKGAVKIDAOSSLTIGSSNMLASAITYSNFGTPEGESIQLDYKDV 356
Oy 116 QIKYQKQLLFKSVNCPGL--TLNSAHN-----CNKNAASGASLYLYP-----AG 160
Db 357 TIGGQIATTTFTFNAPSLTITINSLSKISGDPYSVANPDELIGINFFSVSSGKGDIAG 416
Oy 161 ELKMLPFG--GIMD-----ATKLIRYKRRISSEYGYTITITKLDKGNIQIWL 209
Db 417 KINNIITIGLDVFNFTVASGSGAGGNLPLELENLTIKGGASLGSSTIRSGGGGAV----- 471

```

```

QY 210 OFKSDARVDUN-----LRPT-----GGTYIGRNSVDMCFYDGYSTNSSLE----- 251
Db 472 -FIKSONIDISGOSALLRPSNITSTFGHNGNIDINTLNLITISGGISSSTLSAKA 530
QY 252 -----IRFODNNPK--SDGKPYLRKINDPTKEIAYTLLSLAGK----- 289
Db 531 GNISINSNSINVTGNINSNSPSFINSNELL--VDPNLQKLLYRQPPLLIGQACNIFL 588
QY 290 -----LFTNGTSLN-----IADAASLETMMNRITAVTMEISVPLCWP-GRQLDAKVE 339
Db 589 NTDIINISNGGLINARNEGVDAGNIRISANTININSOGEVNATTTIGEGNIIILNSR-- 646
QY 340 NPEAGQYMGNIIVTFT 355
Db 647 -----NLELNSRITAT 658

```

Search completed: August 19, 2003, 09:49:00
 Job time : 46 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 2003, 09:37:09 ; Search time 23 Seconds

(without alignments)
738.115 Million cell updates/sec

Title: US-09-839-894-10
Perfect score: 1886
Sequence: 1 MNKILFIPLFFSSVLTFA.....EAGYKGNINVTTPSSQTL 361

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	1558.5	82.6	360	1 CFAE_ECOLI	P25734 escherichia
2	120	6.4	2660	1 YEEL_ECO57	Q8XBV7 escherichia
3	114	6.0	650	1 SAGI_YEAST	P20840 saccharomyc
4	109	5.8	2358	1 YEEL_ECOLI	P76347 escherichia
5	108	5.7	524	1 GAG3_YEAST	Q03655 saccharomyc
6	107	5.7	304	1 YDEQ_ECOLI	P77588 escherichia
7	100	5.3	737	1 ALYS_ENTRA	P37710 enterococcu
8	99	5.2	802	1 SYFB_LISIN	Q92C16 listeria in
9	99	5.2	1849	1 IGAI_HAEIN	P45386 haemophilus
10	98.5	5.2	608	1 YD93_METJA	Q58788 methanococ
11	98	5.2	758	1 SP21_YEAST	P35209 saccharomyc
12	98	5.2	1157	1 N133_YEAST	P31855 paenibacill
13	95.5	5.1	713	1 CDG2_PAEWA	P36161 saccharomyc
14	95	5.0	413	1 RF1M_YEAST	P30775 saccharomyc
15	95	5.0	2334	1 WAPA_BACSU	Q07633 bacillus su
16	94.5	5.0	959	1 N100_YEAST	Q07629 saccharomyc
17	94	5.0	471	1 LEU2_BUCRP	P48573 buchnera ap
18	93.5	5.0	412	1 THBG_SHEEP	P50450 ovine arter
19	93	4.9	1167	1 CLAA_BACTU	P56556 bacillus th
20	92.5	4.9	320	1 CYF_CIVCA	Q9T1S4 cyanidulum c
21	92.5	4.9	544	1 AGM1_CANAL	Q9P4V2 candida alb
22	92.5	4.9	1061	1 OAR_MYYXA	P38370 myxococcus
23	92.5	4.9	2278	1 FABL_YEAST	P34756 saccharomyc
24	91.5	4.9	500	1 SYK_BUCBP	Q89655 buchnera ap
25	91.5	4.9	869	1 CFAE_ECOLI	P25733 escherichia
26	91.5	4.9	1783	1 Y468_MYCGE	Q49460 mycoplasma
27	91	4.8	802	1 SYFB_LISMO	Q8Y9Q1 listeria mo
28	90.5	4.8	309	1 ELTB_CLOPE	P30654 xenopus lae
29	90.5	4.8	863	1 MCMA_XENLA	P38073 homo sapien
30	90.5	4.8	1019	1 ENTK_HUMAN	Q09811 schizosacch
31	90.5	4.8	1328	1 HUS2_SCHPO	Q66479 h genome po
32	90.5	4.8	2193	1 POU2_HE71M	P40578 saccharomyc
33	89.5	4.7	1113	1 MGA2_YEAST	

34	89	4.7	465	1 ANT3_MOUSE	P32261 mus musculu
35	88.5	4.7	455	1 MURE_BUCAI	P57315 buchnera ap
36	88	4.7	1394	1 HAP_HAEIN	P45387 haemophilus
37	87.5	4.6	468	1 VGIC_HSEVB	P12889 equine herp
38	87.5	4.6	792	1 UBPA_YEAST	P53874 saccharomyc
39	87	4.6	711	1 ETP2_SFVKA	Q9Q8Y2 Shope fibro
40	87	4.6	1290	1 BXC1_CLOBO	P18640 clostridium
41	86.5	4.6	411	1 THBG_BOVIN	Q9T136 bos taurus
42	86.5	4.6	1002	1 SPHR_AMEPV	P29815 asacta moo
43	86.5	4.6	1273	1 MEB1_YEAST	P38968 saccharomyc
44	86.5	4.6	1367	1 RML_DROME	Q9V7H4 drosophila
45	86.5	4.6	1769	1 YJY9_YEAST	P42945 saccharomyc

ALIGNMENTS

RESULT 1	ID	CFAE_ECOLI	STANDARD;	PRT;	360 AA.
AC	P25734:				
DT	01-MAY-1992 (Rel. 22, Created)				
DT	01-MAY-1992 (Rel. 22, Last sequence update)				
DT	01-AUG-1992 (Rel. 23, Last annotation update)				
DE	CFA/I fimbrial subunit E (Colonization factor antigen I subunit E).				
GN	CFAE.				
OS	Escherichia coli.				
OG	Plasmid NTP513.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
OC	Enterobacteriaceae; Escherichia.				
OX	NCBI_TaxID=562;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRATIN-Enterotoxigenic;				
RX	MEDLINE=69350163; PubMed=2569152;				
RA	Hammers A.M., Pel H.J., Wallshaw G.A., Kusters J.G.,				
RA	van der Zeijst B.A.M., Gastra W.,				
RT	"The nucleotide sequence of the first two genes of the CFA/I fimbrial				
RT	operon of human enterotoxigenic Escherichia coli.";				
RL	Microb. Pathog. 6:297-309(1989).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=92329981; PubMed=1352712;				
RA	Jordi B.J.A.M., Wallshaw G.A., Van der Zeijst B.A.M., Gastra W.;				
RT	"The complete nucleotide sequence of region 1 of the CFA/I fimbrial				
RT	operon of human enterotoxigenic Escherichia coli.";				
RL	DNA Seq. 2:257-263(1992).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL: M55661; AAC41417.1; -				
DR	PIR: D56617; D56617.				
KW	Antigen; Fimbria; Plasmid.				
SO	SEQUENCE 360 AA; 39903 MW; 691509B63A8E9CE CRC64;				
QY	Query Match	82.6%; Score 1558.5; DB 1; Length 360;			
QY	Best Local Similarity	81.2%; Pred. No. 1.6e-112;			
QY	Matches 293; Conservative 31; Mismatches 36; Indels 1; Gaps 1;				
DB	1 MNKILFIPLFFSSVLTFAVADKIPGDESTINIFGPRDRNSSKHNILNNHTAYSE 60				
DB	1 MNKILFIPLFFSSVLTFAVADKIPGDESTINIFGPRDRNSSKHNILNNHTAYSE 60				
QY	SHLTLYRMFELCSSNHTNGACPTSENSSSVSGSETNTLOFTEKRSILKEQLIKY 120				
QY	SHLTLYRMFELCSSNHTNGACPTSENSSSVSGSETNTLOFTEKRSILKEQLIKY 120				
DB	SHLTLYRMFELCSSNHTNGACPTSENSSSVSGSETNTLOFTEKRSILKEQLIKY 120				

Query Match	Similarity	Score	DB	Length		
Best Local	20.98	114	1	650		
Matches	98	Conservative	52	Mismatches 147; Indels 172; Gaps 23		
QY	4	ILFFTFEFSVL	FTFA	VSADKIPG	DESITNIFGRDNE	43
Db	8	ILMFSLALSAININDTFESNLETTPLTANKQPOGWTATPFDESIADASSIREGDEFTL				67
QY	44	SSP	KHNILNNHITA	YSESHTLYDRMFTLC	LSHNTLN	80
Db	68	SMPIVYRIKILINSQOTATISLADGTFAKCYVSOQAALYENTPTCTAONDLSSTYNTID				127
QY	81	GACPTSENPSSVSGETNITLQTFEKSILKRELIQVGNQLEFKSVNC	PSGLTMSA			139
Db	128	GSTIFSLNDSGGSYEYL	ENAKFKRSGPMVLKVLGNQMSDVYNFDPAAETENVF			182
QY	140	HFNCKNAASGASLYL	YPAGELKNLPFGGIDATLKLKRYKRYSETGYTYNITIK			197
Db	183	HSGRSTGSGPESHLMGWCNPGVF	LG	TEKI	DYDSSNNVDL	225
QY	198	LTDKNGIOI	WLPQFSDAVDLNLRTGGGYIQRN	SVMCFPDG		242
Db	226	DCSSVOYVSSNDFNDMWFMPOSYNDTNADV	TEFGSNLMTLDEKLYDGEML			275
QY	243		YSTNSSSLE	IRFODNN	PKSDGEF	265
Db	276	WVNNLQSLPANVNTIDHLEFQVYCLDITANTYTATQSTRTREFIVYGRNLTGSAKSS				335
QY	266	YLRKINDTKET	AYTLS	LLIAGSLPNTGSLNIADAA		304
Db	336	FISTTTDILASINTSAVSTGSI	STVENGNRTTSEVISHVYTTSTKLSTLATSTSLTIAOTS			395
QY	305	SLFTNMN		RITAVIMPELSVVLCPGRL		332
Db	396	IYSDSNITVGTDIHTTSEVIVSDVETISREKTA	STAVVAAPSTTGWTGAM			444

ID	YEEL	ECOLI	STANDARD:	PRT:	2358 AA.
AC	P76347		p94750;		
DT	28-FEB-2003	(Rel. 41, Created)			
DT	28-FEB-2003	(Rel. 41, Last sequence update)			
DT	28-FEB-2003	(Rel. 41, Last annotation update)			
DE	Hypothetical protein yeef.				
GN	YEEL OR B1978.				
OS	Escherichia coli.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
OC	Enterobacteriaceae; Escherichia.				
OX	NCBI_taxonomy:562;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-K12 / MG1655;				
RX	MEDLINE=97426617; PubMed=9278503;				
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,				
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,				
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,				
RA	Mau B., Shao Y.;				
RA	"The complete genome sequence of Escherichia coli K-12.";				
RA	Science 277:1453-1474(1997).				
RA	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-K12;				
RX	MEDLINE=97251358; PubMed=9097040;				
RA	Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,				
RA	Iseno K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,				
RA	Makino K., Maki T., Mizobuchi K., Mori H., Mori T., Motomura K.,				
RA	Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T.,				
RA	Saito N., Sampei G., Seki Y., Sivasubraman S., Tagami H.,				
RA	Takeda J., Takemoto K., Wada C., Yamamoto Y., Horuchi T.;				
RA	"A 460-kb DNA sequence of the Escherichia coli K-12 genome				
RT	corresponding to the 40.1-50.0 min region on the linkage map.";				
RT	DNA Res. 3:379-392(1996).				
RL	-1 SIMILARITY: Contains 13 big-1 domains.				
CC	-1 SIMILARITY: BELONGS TO THE INTIMIN/INVASIN FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyrighted. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL; AE000289; AAC75042.1; ALT_INIT.				
CC	EMBL; D90837; BAAL5800.1; -				
CC	EMBL; D90836; BAAL5799.1; ALT_INIT.				
CC	Ecogene; EG13378; yeef.				
DR	InterPro; IPR003344; Big_1.				
DR	InterPro; IPR003533; Intimin.				
DR	InterPro; IPR002482; LysM.				
DR	InterPro; IPR000601; PKD_domain.				
DR	Pfam; PF02369; Big_1; 13.				
DR	PRINTS; PRO1369; INTIMIN.				
DR	SMART; SMO0634; BID_1; 13.				
DR	SMART; SMO0257; LysM; 1.				
DR	SMART; SMO0089; PKD; 6.				
KW	Hypothetical protein; Repeat; Complete proteome.				
FT	DOMAIN 738 834		BIG-1 1.		
FT	DOMAIN 840 931		BIG-1 2.		
FT	DOMAIN 932 1033		BIG-1 3.		
FT	DOMAIN 1042 1137		BIG-1 4.		
FT	DOMAIN 1146 1237		BIG-1 5.		
FT	DOMAIN 1246 1350		BIG-1 6.		
FT	DOMAIN 1351 1448		BIG-1 7.		
FT	DOMAIN 1449 1553		BIG-1 8.		
FT	DOMAIN 1554 1655		BIG-1 9.		
FT	DOMAIN 1661 1754		BIG-1 10.		
FT	DOMAIN 1763 1853		BIG-1 11.		
FT	DOMAIN 1855 1950		BIG-1 12.		
FT	DOMAIN 1952 2053		BIG-1 13.		

FT	CONFLICT	105	105	S -> G (IN REF. 2).
SO	SEQUENCE	2358 AA;	248599 MM;	232249750BF631ED CRC64;
QY	Query Match	5.8%;	Score 109;	DB 1; Length 2358;
	Best Local Similarity	20.0%;	Pred. No. 3.7;	Mismatches 157; Indels 90; Gaps 15;
Dd	Matches	78;	Conservative 65;	
OY	13	SSVLETFEVSADKIPGDESTINIFGPRDRNESSPKHNTLNHHITAY---	SESHLYDGMT 69	
	: : : : : : : : : : : : : :			
Dd	737	SAKITATLSANNGVLNENANAITYSVNVADSGS---NINDHTYTFVLLSSAITSFNQN	793	
OY	70	FLCJSSHNLTACAPCTSENPSSSVSGETNTLQTEKRSLI-----	K 112	
	: : : : : : : : : : : : : :			
Dd	794	----TAKTDVGNGLA-TFDLKSSKOEDNTVEVTELVENGWYKQTILIVSFVGDSSTAQVDLOKSK	848	
OY	113	RELQIKGYKCQ-LRKSVCNPGCLTLNSHFNCNNNAASGALYLITPAGELKNLPFGIW	171	
	: : : : : : : : : : : : : :			
Dd	849	NEVVADGGDVSTMTATVAVDAKGNLLNDVMYTEFVNNSAAKLSTQEIVNSHD-----	GIA 901	
OY	172	DATKLRKYRRYSSETGYTYTNTITKLDKONIOIMLPQFSRDARVDNLPTGGGYIG	231	
	: : : : : : : : : : : : : :			
Dd	902	TATLT-----SLKNGDYRTASVSSGSQAQQVNFIDGSTAALTISTV-PSGDITV--	951	
OY	232	RNSVDMCFYDGSTNSS---SLEIRQDNPPKSDGFYLRKINDDTKEIALVTLSLLAG	287	
	: : : : : : : : : : : : : :			
Dd	952	-----TNTAPQYMTATLQDKNGNPDLKETTESVPND---VASKSPISNGG	994	
OY	288	KSLEPTNGTSL-----NINDAASLETNRMTTRAVTPMETSVPVLC	327	
	: : : : : : : : : : : : : :			
Dd	995	KGMTDMSNCVAIASLTGLTAGTHMMARIANSNSVSDAQPMITPVADKRAVYVLOTSKAEIT	1054	
OY	328	WPG--RLQLDAKVENPEAGQYMGMINVETF	355	
	: : : : : : : : : : : : : :			
Dd	1055	NGVDETTLTATVKDP-SNHFPVAGITVFET	1083	
RESULT 5				
GAS3_YEAST	STANDARD:	PRT;	524 AA.	
ID	GAS3_YEAST			
AC	Q003655;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	GAS3 protein precursor.			
GN	GAS3 OR YMR215W OR YM8261.09.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
CC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
OX	NCHI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288c / AB972;			
RX	PubMed=9169872;			
RA	Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,			
RA	Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,			
RA	Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,			
RA	Rice P., Skelton J., Walsh S., Whitehead S., Barrall B.C.,			
RT	"The nucleotide sequence of Saccharomyces cerevisiae chromosome			
RT	XIII."			
RL	Nature 387:90-93(1997).			
RN	{2}			
RP	IDENTIFICATION.			
RX	MEDLINE=20529944; PubMed=11079560;			
RA	Pardo M., Ward M., Bains S., Molina M., Blackstock W., Gil C.,			
RA	Nombela C.;			
RT	"A proteomic approach for the study of Saccharomyces cerevisiae cell			
RT	wall biogenesis."			
RL	Electrophoresis 21:3396-3410(2000).			
CC	-I- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor			
CC	(potential)			
CC	-I- SIMILARITY: Belongs to the GAS1 family.			
CC	-----			
CC	This SWISS-PROT entry is copyright It is produced through a collaboration			

DT 15-SEP-2003 (rel. 42, last annotation update)
 DE Autolysin precursor (EC 3.2.1.-) (Peptidoglycan hydrolase)
 DE (Beta-glycosidase).
 GN EF0799.
 OS Enterococcus faecalis (Streptococcus faecalis).
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
 OX NCBI_TaxID=1351;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91358349; PubMed=1679432;
 RA Bellevue C., Potvin C., Trudel J., Asselin A., Bellemare G.;
 RT "Cloning, sequencing, and expression in *Escherichia coli* of a
 RL Streptococcus faecalis autolysin";
 RL J. Bacteriol. 173:5619-5623(1991).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=V583 / ATCC 700802;
 RX MEDLINE=2250657; PubMed=12663927;
 RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,
 Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
 Tettelin H., Dodson R.J., Umayam L., Brinkac L., Beanan M.,
 Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Madupu R., Nelson W.,
 Vamathevan J., Tran B., Upton J., Hansen T., Shetty J., Khouri H.,
 Ueberbach T., Radue D., Ketchum K.A., Dougherty B.A., Fraser C.M.;
 RA "Role of mobile DNA in the evolution of vancomycin-resistant
 RT Enterococcus faecalis";
 RL Science 299:2071-2074(2003).
 CC -1- FUNCTION: Hydrolyzes the cell wall of *E. faecalis* and
 CC M. lysodeikticus. May play an important role in cell wall growth
 CC and cell separation.
 CC -1- SUBCELLULAR LOCATION: Secreted (Probable).
 CC -1- DOMAIN: LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOGLYCAN
 CC BINDING.
 CC -1- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.
 CC -1- SIMILARITY: Contains 6 LysM repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M58002; AAA67325.1; -
 DR EMBL: AE016949; AA080613.1; -
 DR PIR: A38109; A38109.
 TIGR: EF0799; -
 Interpro: IPR002901; Amidase_4.
 DR Interpro: IPR002482; LysM.
 DR Pfam: PF01832; Amidase_4; 1.
 DR Pfam: PF01476; LysM; 5.
 DR SMART: SM00257; LysM; 6.
 DR SMART: SM00047; Lys2; 1.
 DR Hydrolase: Glycosidase; Bacteriolytic enzyme; Cell wall;
 KM Cell division; Septation; Repeat; Signal.
 FT SIGNAL 1 53
 FT CHAIN 54 737
 FT REPEAT 363 405
 FT REPEAT 431 473
 FT REPEAT 499 541
 FT REPEAT 567 609
 FT REPEAT 633 675
 FT REPEAT 695 737
 FT REPEAT 85 85
 FT CONFLICT 118 118
 FT CONFLICT 143 143
 FT CONFLICT 417 417
 FT CONFLICT 449 449
 FT CONFLICT 476 476
 FT CONFLICT 484 484
 FT CONFLICT 567 632
 FT SEQUENCE 737 AA; 77025 MW; ABB16BD506AC7507 CRC64;

Query Match 5.3%; Score 100; DB 1; Length 737;
 Best local Similarity 22.9%; Pred. No. 4.1;
 Matches 69; Conservative 39; Mismatches 115; Indels 78; Gaps 15;
 QY 41 RNESSPRHNI-LNNHTAVSESHLYDRMTFLCLSHNTLNGACPTSPNSSSV----- 94
 DB 309 RYATDPSPYAKLNINVTAY--NLTYDTPSSGNTGGTGVNPGTSGNNSQSTNYTYVK 366
 QY 95 SGET--NTLQF-----TEKRS--IKREIQIGYQDLFKSWNCSGLTLSNAHFNCKNN 146
 DB 367 SDDTLNKKIAAQYGVSVANLRSMNGISGDLIFVQKLIYKKA--SGNTGGSGGSGNN 423
 QY 147 AASGASLYIPAGELKNLPFGIMDATLKLVRKRSYSETGYTITINILIKLTDGNIQI 206
 DB 424 -QSGNTYTYTVASGDTLN-----KIAAQY-----VSANLR 455
 QY 207 WLQPKSDARVDLNRPTGGGTYIGRNSVDMCFYDGYSTNSSLEIRPDNNPKSDGFY 266
 DB 456 W-----NGISGDL-----IFVGQK--LIVKKGASGNTGSGNNGSGNNSQSTNYTY 499
 QY 267 LKINDPRKEIA--YTLG-----LLAGKSLTPNGSLNIAADASLETNN 311
 DB 500 TTKSGDTLKKIAAQYGVSVANLRSMNGISGDLIFAGQKIYKKGTSNTGSSNGSGNN 559
 QY 312 R 312
 DB 560 Q 560
 RESULT 8
 SYFR_LISIN
 ID SYFR_LISIN STANDARD; PRT; 802 AA.
 AC Q92C16;
 DT 15-SEP-2003 (rel. 42, last annotation update)
 DT 15-SEP-2003 (rel. 42, last annotation update)
 DE Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20)
 DE (Phenylalanyl-tRNA ligase beta chain) (Phers).
 GN PHER OR L1N1185.
 OS *Listeria innocua*.
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
 OX NCBI_TaxID=1642;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CLIP 11262 / Serovar 6a;
 RX MEDLINE=21537279; PubMed=11679669;
 RA Glaser P., Frangoul L., Buchrieser C., Rusniok C., Amend A.,
 Baquero F., Berche P., Bloeker H., Brandt P., Charrabory T.,
 Charbit A., Chelouani F., Couve E., de Baurvar A., Dehoux P.,
 Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussuguet O.,
 Entlan K.-D., Eshti H., Garcia-del Portillo F., Garrido P.,
 Gautier L., Goebel W., Gomez-Lopez N., Hain T., Haut J., Jackson D.,
 Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurupkat G.,
 Madueno E., Maitouram A., Mata Vicente J., Ng E., Nedjati H.,
 Nordstiek G., Novella S., de Padlos B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schlueter T., Simoes N., Tietz A.,
 RA Vazquez-Boland J.-A., Voss H., Wehlund J., Cossart P.;
 RT "Comparative genomics of *Listeria* species";
 RL Science 294:849-852(2001).
 CC -1- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) - AMP +
 CC diphosphate + L-phenylalanyl-tRNA(Phe).
 CC -1- COFACTOR: Binds 2 magnesium ions per tetramer (By similarity).
 CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (By
 CC similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to the phenylalanyl-tRNA synthetase beta chain
 CC family. Subfamily 1.
 CC -1- SIMILARITY: Contains 1 tRNA-binding domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its


```

Db      954 TLFDSNATRNNEVTLANGSVDRGAMKKYLRNVNG--RYDLYNEVEKRNQTVDTNIT 1011
QY      297 SLN--IADAASLETNMMRTAVTMP 319
      : | | | : | | |
Db      1012 TPNDIQADAPSAQSNNEETARVETP 1036

RESULT 10
ID      YD93_METUA          STANDARD:          PRT:          608 AA.
AC      058788:
DT      15-JUL-1998 (Rel. 36, Created)
DT      15-JUL-1998 (Rel. 36, Last sequence update)
DE      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Hypothetical protein MJ1393.
GN      MJ1393.
OS      Methanococcus jannaschii.
OC      Archaea: Euryarchaeota; Methanococci; Methanococcales;
OC      Methanocaldococcaceae; Methanocaldococcus.
NCBI_taxid=2190;
(1)
RP      SEQUENCE FROM N.A.
RC      STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX      MEDLINE=96337999; PubMed=8688087;
RA      Bulc C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA      Sutton G.G., Blake O., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA      Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA      Overbeek R., Kirkness E.F., Weissman K.G., Merrick J.M., Glodek A.,
RA      Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhman J.L., Nguyen D.,
RA      Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA      Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA      Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT      "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT      jannaschii."
RL      Science 273:1058-1073(1996).
CC      -I SIMILARITY: TO M.JANNASCHII MJ1394 AND A.FUIGIDUS AF2028.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: U67579; AAB99403.1; -.
DR      PIR: H64473; H64473.
DR      TIGR: MJ1393; -.
FT      TRANSMEM 4 24 POTENTIAL
SQ      SEQUENCE 608 AA: 66768 MW: 010FAFLC29F8C73C CRC64;

Query Match
Best Local Similarity 20.3%; Pred. No. 4.1;
Matches 85; Conservative 68; Mismatches 119; Indels 147; Gaps 25;

QY      1 MAKILF---ITFLFFSVLFTFAVSADKP-----GDESTINITEPRDR 41
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      1 MKRLFLMALMSLFLTGTVGY--GDNGPLVAVYEKYNINGNTGDLVSTI----- 52
QY      42 NESSPKHNLNNHTTAYSESHLYDRMTFLCISNHT-----LNG-----ACPISE 87
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      53 -ESTIGTVYINN--TGTITNDLYDVWVAVNISNNITGPEYVNGTPKGVFISSAPAVT 109
QY      88 N-PSSS-----VSGETNITLQFTKRSLLKRELOIGYKQLLFKSVNCPGLTLN-SA 139
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      110 NLPNANTYIHIPILPNNSVYIKFAIDKSIICVPLINE-----TSDKIRIPSEKISNMSV 165
QY      140 HPNCKNAASGASLYLYPAGE-----LKNLPFGGIWDATLKLVRKRYSETYGT 189
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      166 YLINSIRNSVA-----LPATDTPVSVIMTKLISNDP-----NNYGS 200
QY      190 YT---INTIKLTGKNGIQI---LPQFK-----SDARVDLNLRTPGGCT 228

```

```

Db      201 DTWNLNITGALANGSITIMDGPYFLPGYNDSLTWGVIINTKNAITITINI--TGNN 258
QY      229 YIGRSVDMCFYDGTSTNSSSLIEIFQDNNKPSDDKFLRKINDTKRIATY-----LSL 283
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      259 YTNRTGTLMKY--GFAV-----IFFEENGTKSGTKI-----EGIATGAGYVSA 300
QY      284 ILAGSLTPITNG-----TSLNIADAASLETNMMRTAVTMPDI--SVPLQWPGRLQ 334
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      301 TREGPFLNASSCKYIWEANVSKAS--SYFNLTHVTIMAVNGSPVILDPENITL 357

RESULT 11
ID      SP21_YEAST          STANDARD:          PRT:          758 AA.
AC      P35209:
DT      01-FEB-1994 (Rel. 28, Created)
DT      01-FEB-1994 (Rel. 28, Last sequence update)
DE      15-SEP-2003 (Rel. 42, Last annotation update)
DE      SP21 protein.
GN      SP21 OR YMR179W OR YMR8010.09.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomyces.
NCBI_taxid=4932;
(1)
RP      SEQUENCE FROM N.A.
RC      MEDLINE=9418609; PubMed=8138180;
RX      Natsoulis G., Winston F., Boeke J.D.;
RA      "The SP21 and SP22 genes of Saccharomyces cerevisiae."
RT      Genetics 136:93-105(1994).
RL      [2]
RN      -----
RN      SEQUENCE FROM N.A.
RC      STRAIN=S288C / AB972;
RX      PubMed=9169872;
RA      Bowman S., Churche C.M., Badcock K., Brown D., Chillingworth T.,
RA      Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA      Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA      Rice P., Skelton J., Walsh S., Whitehead S., Barrett B.G.;
RT      "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT      XIII."
RL      Nature 387:90-93(1997).
CC      -I FUNCTION: REQUIRED FOR NORMAL TRANSCRIPTION AT A NUMBER OF LOCI IN
CC      YEAST.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: L24436; AAA35078.1; -.
DR      EMBL: Z49808; CAA89912.1; -.
DR      PIR: S47866; S47866.
DR      TRANSEC: T04376; -.
DR      SGD: S0004791; SP21.
DR      GO: GO:0006357; P:regulation of transcription from pol II pro. .; IMP.
FT      DOMAIN 127 144 ASP/GLU-RICH (ACIDIC).
FT      DOMAIN 672 682 ASP/GLU-RICH (ACIDIC).
SQ      SEQUENCE 758 AA: 84697 MW: 7DB3FCF7EB96705 CRC64;

Query Match
Best Local Similarity 21.9%; Pred. No. 6;
Matches 61; Conservative 34; Mismatches 95; Indels 88; Gaps 12;

QY      42 NESSPKHNLNNHTTAYSESHLYDRMTF-----LCISSNLTNGACPTSENPSSSSVS 95
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      166 NISNKKGRVNNQI-----PEETLEVKLRTFYITLRLFSGNNTNSRISCLQMPSSSL--- 218
QY      96 GETNITLQFTKRSLLKRELOIGYKQLLFKSVNCPGSLTINSASHFNCKNAASGLYL 155
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

Db 219 --PSATLPEPTPKSOSLEKTNQIR-----NSRNARTITIN-----NTNSGT-----257
 QY 156 YIPAGELKNLPFGCINPATLTKRYKRRYSERYGYTINTITIKLTKDKNQIOWLPQFSDA 215
 Db 258 ---VGRKQTPM---PAPAVRQ-----SLPIW-----280
 QY 216 RVDLNLPTGGGYTIGNSVDMCFY-----DGSTNSSSLERFODNNRPSDKFYLRK 269
 Db 281 ----NLKPNLANTGFPNRSIAHKIYLDRKTEANQONQONHONTAYELINTLQNDNTIQRTK 336
 QY 270 INDD--TKETAYTISLLAGKSLPTNCT-----SLNI 300
 Db 337 IDDSVSKRFDFMLNKRKSTKVKSPGATIAKKPASINI 374

RESULT 12

ID N133_YEAST STANDARD; PRT; 1157 AA.
 AC P36161;
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 GN Nucleoporin NUP133 (Nuclear pore protein NUP133).
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JUXJR;
 RX MEDLINE=95112817; Pubmed=7813444;
 RA Doye V., Wept R., Hurt E.C.;
 RT "A novel nuclear pore protein Nup133p with distinct roles in poly(A)+
 RNA transport and nuclear pore distribution.";
 RL RNA J. 13:6062-6075(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9426327; Pubmed=8203164;
 RA Garcia-Cantalejo J., Baladron V., Esteban P.F., Santos M.A., Bou G.,
 Ramacia M.A., Revuelta J.L., Ballesta J.P.G., Jimenez A., del Rey F.;
 RT "The complete sequence of an 18,002 bp segment of Saccharomyces
 cerevisiae chromosome XI contains the HBS1, MRP-120 and PRP16 genes,
 and six new open reading frames.";
 RL Yeast 10:231-245(1994).
 CC -1- FUNCTION: INVOLVED IN POLY(A)+ RNA TRANSPORT AND NUCLEAR PORE
 DISTRIBUTION.
 CC -1- SUBCELLULAR LOCATION: Nuclear pore complex.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL: X80066; CA56372.1; -;
 DR EMBL: Z27116; CA81633.1; -;
 DR EMBL: Z28307; CA82161.1; -;
 DR PIR: S38160; S38160.
 DR SGD: S0001790; NUP133.
 DR GO: GO:0005643; C:nuclear pore; IDA.
 DR GO: GO:0006406; P:mRNA-nucleus export; IMP.
 DR GO: GO:0006999; P:nuclear pore organization and biogenesis; IMP.
 DR Pfam: PF04044; Nup133; 1.
 DR Nuclear protein; Transport.
 SQ SEQUENCE 1157 AA; 133319 MW; C8BDBB7D709C5C08 CRC64;

Query Match 5.28; Score 98; DB 1; Length 1157;
 Best Local Similarity 22.18; Prod. NO. 10;
 Matches 75; Conservative 43; Mismatches 127; Indels 94; Gaps 21;

QY 48 HNLNHNITAY-----SESHLYDRMTF-----LCLSHNHTNGACPTSENP 89
 Db 96 YGLVNDHKRYIWNHISHTQDTPYITVFPSSDNDDEIAVAPRCILFEPATWDESLPALNP 155
 QY 90 SSSSVSGE-----TNITLQTEKSLKRELQIR-----GYQLLFK 126
 Db 156 NDDDETGLIINKSKAIYEDINSINNLFKLEKFS---HELELPINSSGGGKCDLM- 211
 QY 127 SYNC-PSGLNLNS-----AHNCKNKAASGASLYIYIPAGELKNLPFG-GLMDATLKLRL- 178
 Db 212 -LNCPEAGIVLSTMGRIFFITJINSKGFQQLK-GLLNLPKLGIMSKIFNTNSS 266
 QY 179 -VKRRYSEFY--GYTINTITIKLTKDKNQIOWLPQFSDARVDLNLPTGGGYTIGNSV 235
 Db 267 VVSLRNGPIILGKGRLYIIT--TKGIRPTW--QLSA-----TNSHPT-----KLI 308
 QY 236 DMCFYDGYSTNSSSL-----ELRFODNNRPSD--GKFIYRKINDT-KEIAYTISLLA 286
 Db 309 DVNIYEALIESLDLPFPFAGTCLKIMDSHPLODESSQLFLSYIDSSCNFTYILSTYIF 368
 QY 287 GKSLFTPTNGTSLNADAASLETNNRITAYTM-PEISVP 324
 Db 369 DSS-----SNSFTLFTYRLNTEMSITDTRKFRKIRIP 402

RESULT 13

CDG2_PAEMA
 ID CDG2_PAEMA STANDARD; PRT; 713 AA.
 AC P31835;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Cyclomalodextrin glucanotransferase precursor (EC 2.4.1.19)
 DE (Cyclodextrin-glycosyltransferase) (Cgtase).
 OS Paenibacillus macerans (Bacillus macerans).
 OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
 OX NCBI_TaxID=44252;
 RN [1]
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 28-37.
 RA Sugimoto T., Kubota M., Sakai S.;
 RT "Polypeptide possessing cyclomalodextrin glucanotransferase
 activity.";
 RL Patent number GB2169902, 23-JUL-1986.
 CC -1- CATALYTIC ACTIVITY: Degrades starch to cyclodextrins by formation
 of a 1,4-alpha-D-glucosidic bond.
 CC -1- COFACTOR: BINDS TWO CALCIUM IONS.
 CC -1- SUBUNIT: Monomer.

CC -1- MISCELLANEOUS: CGTASE MAY CONSIST OF TWO PROTEIN DOMAINS: THE ONE
 IN THE AMINO-TERMINAL SIDE CLEAVES THE ALPHA-1,4-GLUCOSIDIC BOND
 IN STARCH, AND THE OTHER IN THE C-TERMINAL SIDE CATALYZES OTHER
 ACTIVITIES, INCLUDING THE RECONSTITUTION OF AN
 ALPHA-1,4-GLUCOSIDIC LINKAGE FOR CYCLIZING THE
 MULTIOLOGOSACCHARIDE PRODUCED.
 CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
 KNOWN AS THE ALPHA-AMYLASE FAMILY.

CC PIR: S26589; ALBSXR.
 DR HSSP: P43379; ICDG.
 DR InterPro: IPR006589; Alp_amy1_cat_sub.
 DR InterPro: IPR006048; Alpha_amy1_C.
 DR InterPro: IPR006047; Alpha_amy1_cat.
 DR InterPro: IPR002044; CBD_4.
 DR InterPro: IPR006046; Glyco_hydro_13.
 DR InterPro: IPR002909; IPT_TIG.
 DR Pfam: PF00128; alpha-amy1ase; 1.
 DR Pfam: PF02806; alpha-amy1ase_C; 1.
 DR Pfam: PF00686; CBM_20; 1.
 DR Pfam: PF01833; TIG; 1.
 DR PRINTS: PR00110; ALPHAMYLASE.
 DR PRODOM: PD001568; CBD_4; 1.
 DR SMART: SM00642; Amyy; 1.
 DR SMART: SM00632; Amyy_C; 1.
 DR Transferrase; Glycosyltransferase; Calcium; Signal.
 FT SIGNAL 1 27


```

QY 135 TINSAHFNCNKNASGASLYIYPAGELKNLPFGIMDAFLKLRVRRSEYGYTTINI 194
Db 347 VLKTYGYY---DKTTGTN-YAFMKFNNLKPIONMTYTKATLKTYVAHSY---YGT----- 393
QY 195 TIKLFDKGNIOIWLPOFKS---DARVDNLNRPYGGTYIGRNSVDMCFYDGY-----S 244
Db 394 --KATG-----LMLDTVNSNYDNNAKYTWNTKFPASKN--IGKADYHKQWNASYDYTAAVKS 444
QY 245 TNSSSLEIRFQDNNPKSDGKFFYLKRI-----NDDTKEIAVTLISLLAGKSLTPT----- 293
Db 445 WNSGANYGFK-LHTNGNGKEYWKLISSANSANKRPIEYTYTIP-----KGNTPITKAY 498
QY 294 -NCTSLINADASLETNNMRI 313
Db 499 HNGDSTGYFD-----ISMKKV 514

```

Search completed: August 19, 2003, 09:46:26
 Elapsed time : 29 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 2003, 09:43:34 ; Search time 99 Seconds

(without alignments)
940,980 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1886
Sequence: 1 MNKILFTLFFSSVLFTFA.....EAGQYMGINVTFTPSQTL 361

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: SPREMBL_23:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeal:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1886	100.0	361	2	Q93G67 escherichia
2	949.5	50.3	363	2	Q47115 escherichia
3	930.5	49.3	364	2	Q47119 escherichia
4	381	20.2	387	2	Q8KQ77 burkholderi
5	371	19.7	387	2	Q8KQ10 burkholderi
6	275.5	14.6	359	16	Q9XDS1 salmonella
7	127	6.7	1881	16	Q8RCK2 fuscobacteri
8	115.5	6.1	700	2	Q8VU50 chlamydia p
9	115.5	6.1	847	2	P71132 chlamydomoni
10	114	6.0	865	15	Q8G7H7 human immun
11	113	6.0	2646	5	Q81521 plasmodium
12	111.5	5.9	743	16	Q8XNM2 clostridium
13	111	5.9	2768	16	Q8E9G6 shewanella
14	110.5	5.9	589	16	Q8F994 leptospira
15	110.5	5.9	1052	17	Q8Q0S0 methanosa
16	110	5.8	265	2	Q8KGS5 rhizobium 1

17	110	5.8	846	2	P71133 chlamydomoni
18	109.5	5.8	2454	5	Q8T2G3 dictyosteli
19	109	5.8	807	5	Q8T7V5 drosophila
20	108	5.7	872	2	Q9RIA0 rickettsia
21	108	5.7	2529	16	Q25579 helicobacte
22	107	5.7	304	16	Q8XAX2 escherichia
23	106.5	5.6	1012	16	Q8PER7 xanthomonas
24	106.5	5.6	1065	16	Q8EXX0 leptospira
25	106	5.6	1937	5	Q8IE94 plasmodium
26	105.5	5.6	430	3	Q96VM2 blumeria gr
27	105.5	5.6	843	16	Q8YK08 anabaena sp
28	105	5.6	598	16	Q8Y7I7 listeria mo
29	104.5	5.5	1441	16	Q9CFL1 lactococcus
30	104	5.5	304	16	Q8CW27 escherichia
31	104	5.5	803	10	Q9AUM5 oryza sativ
32	103.5	5.5	32	16	Q9RZS7 sulfolobus
33	103.5	5.5	1269	17	Q97Z06 delnoccoc
34	103.5	5.5	2399	16	Q9ZKS9 helicobacte
35	103	5.5	868	17	Q8PZM5 methanosa
36	103	5.5	1754	16	Q8RD81 thermococ
37	102.5	5.4	1736	5	Q95PH7 dictyosteli
38	102.5	5.4	2747	2	Q91800 aeromonas s
39	102	5.4	376	5	Q19229 caenorhabd
40	102	5.4	692	17	Q97YM6 sulfolobus
41	101	5.4	881	16	Q8EW44 mycoplasma
42	101	5.4	1238	16	Q8ZJ21 yersinia pe
43	101	5.4	1267	16	Q8CLO7 yersinia pe
44	101	5.4	2082	5	Q812R1 plasmodium
45	100.5	5.3	914	16	Q8EX32 mycoplasma

ALIGNMENTS

RESULT 1

ID Q93G67 PRELIMINARY; PRT: 361 AA.
AC Q93G67;
DT 01-DEC-2001 (TREMBLrel, 19, Created)
DT 01-DEC-2001 (TREMBLrel, 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel, 19, Last annotation update)
DE Minor pilin protein CsaE.
GN CsaE.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Bacteria; Proteobacteria; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El1881A;
RA Altboum Z.D., Levine M.M., Galen J.E., Barry E.M.;
RT "Isolation and characterization of EPEC Csa4 fimbriae encoding genes,
RT and their expression in Shigella flexneri 2a quana strain CVD 1204.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL, AF296132; AAK97137.1; ZIE17AABDB0EB353 CRC64;
SQ SEQUENCE 361 AA; 40102 MW; 2E1E7AABDB0EB353 CRC64;

Query Match 100.0%; Score 1886; DB 2; Length 361;
Best Local Similarity 100.0%; Pred. No. 7.1e-137;
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNKILFTLFFSSVLFTFAVSADKIPGDESTINIFGPRDRNESSPKHNILNHNITAYSE 60
DB 1 MNKILFTLFFSSVLFTFAVSADKIPGDESTINIFGPRDRNESSPKHNILNHNITAYSE 60
QY 61 SHITLDKMTFLCUSHNTLNGACPTSPNPSSSVSGTNTITLQFTKRSIKRELQKGY 120
DB 61 SHITLDKMTFLCUSHNTLNGACPTSPNPSSSVSGTNTITLQFTKRSIKRELQKGY 120
QY 121 KOLLFKSVNCPGSLTLNSAFHNCNKNAASASLYLYIPAGELNKLPPGGIMDTLKLRYK 180
DB 121 KOLLFKSVNCPGSLTLNSAFHNCNKNAASASLYLYIPAGELNKLPPGGIMDTLKLRYK 180

Db		: :	1	353 YITFESVNTL	363
RESULT 3					
ID	047119	PRELIMINARY;	PRT;	364 AA.	
AC	047119:				
Dt	01-NOV-1996 (TREMBLrel. 01, Created)				
Dt	01-NOV-1996 (TREMBLrel. 01, Last sequence update)				
Dt	01-NOV-1998 (TREMBLrel. 08, Last annotation update)				
DE	CodD precursor.				
GN	Escherichia coli.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
CC	Enterobacteriaceae; Escherichia.				
OX	NCBI_TaxID=562;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C91f-6;				
RX	MEDLINE=96071908; PubMed=7591145;				
RA	Froehlich B.J., Karakashian A., Skellariaris H., Scott J.R.:				
RT	"genes for cs2 pili of enterotoxigenic Escherichia coli and their				
RL	interchangeability with those for CS1 pili."				
DR	Infect. Immun. 63:4849-4856(1995).				
KW	EMBL; Z47800; CAA87763.1 - "				
FT	SIGNAL.				
SO	SEQUENCE	1 18 POTENTIAL.			
SC	CHAIN	19 364 CONTD.			
SE	SEQUENCE	364 AA; 40979 MW; EC7AC738A0E195B6 CRC64;			
Query Match		49.3%; Score 930.5; DB 2; Length 364;			
Best Local Similarity		48.4%; Pred. NO. 1.8e-63;			
Matches 177; Conservative		64; Mismatches 118; Indels 7; Gaps 3			
QY	1 MNKILFIETLFESSVLTFEFAVSADKIPGDESITIFGRDRY--ESSPKHNLTNNHTAY	58			
DB	1 MKKIYIIVLSMFLCSGVISQSMHTINVEASINKTESIGPIDRSAASYPAHYTFHEHVAGY	60			
QY	59 SESHLYDMRMTFLCLSSHNTINGACTPSSENBSVVSGENITTLQFTEKRSLIKRELQIK	118			
DB	61 NKDSLEFRMFELCMSSTDASKGACPTGENSKS--OGETNIKILFTEKKSLARRTLMNK	118			
QY	119 GYKDLFPFSVNC---PSGLTINSAHFNKNKMAASGLYIVYPAGEIKMLPPGGIMDATL	175			
DB	119 GYKFPLYSDDPCIRHYDKMNLSHTVKCVGSFTRGVDLTLYIPGEIDGLTGIGWEATL	178			
QY	176 KLRVRRARYSETFYGYTTINIITKLTDKGIQLWLPOFKSDARVDLNLPTGGGTYYIGRSVY	235			
DB	179 ELRFVRKHVDYNHGTYKVNIYDLFDKGNIQVTPFXHSDPRIDLMLPREGNGKYGSNVLY	238			
QY	236 DMCCTYDGSTNSSLSLEIFQDNPNRKSDCKFYLIKINDOTKEALYLLSLILAGKSILTPNG	295			
DB	239 EMCICYLDGSTHSOSTEMRFODDSQTGNNEVYLNIKIGEPKIKLPYLLSLLGGEREYPNNG	298			
QY	296 TSLNIADASLETWNMRITAYTMPERISVPVLCWPGRLDLDAKVENPEAGQYMGNIINVTFT	355			
DB	299 EAFITNDTSLEFIWMNRKYSVLPDISIVLCWPANLFLFMSELNPPEAGEYSGILNVPTT	358			
QY	356 PSSQTL 361				
DB	359 PSSSSL 364				
RESULT 4					
ID	08KOJ7	PRELIMINARY;	PRT;	387 AA.	
AC	08KOJ7:				
Dt	01-OCT-2002 (TREMBLrel. 22, Created)				
Dt	01-OCT-2002 (TREMBLrel. 22, Last sequence update)				
Dt	01-MAR-2003 (TREMBLrel. 23, Last annotation update)				
DE	Putative minor pilin protein.				
GN	CBLD.				

Matches 114; Conservative 50; Mismatches 144; Indels 52; Gaps 13;

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 359 AA; 39741 MW; BB88A4E9A62052CE CRC64;
 Query Match 14.6%; Score 275.5; DB 16; Length 359;
 Best Local Similarity 29.7%; Pred. No. 3.6e-13;
 Matches 90; Conservative 48; Mismatches 130; Indels 35; Gaps 11;
 QY 70 FCLSHNTLNGACPTSENPSSSVSGETNITLQFTEKRSLLKRELQIKGYKQLLEKSVN 129
 Db 79 WVCRSNRNENEGACEETHLWVWYAFGAYSKIRLRFREQISHAEITL-----ILLGSR 131
 QY 130 --CPGSLTINSAHFNCNKNAAS---GASLYIIPAGELKNLPFGGTYWDITLKL-RKRRY 183
 Db 132 DACYGV-----INMNAACQWGRSLKLRIPSEELAKIPSGWKATLVLDYLOMG 183
 QY 184 SETYGYTITITIKLTD--KGNIQIWLPOF-KSDARVDLNLPTGGGTYIGRNSVDMCFY 240
 Db 184 DDEPLGISTDITLANTDHEAENAAITFPQFGATATPRVDNLHRMNASQMSGRANLDMCLX 243
 241 DGTSTNSSLEIRFODNNPKSDGKEYLRKINDTKEIATYLSLLAGKSLPTNGTSLNI 300
 244 DG-GVKARSLQMKIEGSKNGSGTGFOYIKSDSADT--IDYAVSMNYGSRISIPVTRGYEFL 300
 QY 301 ADAASLETMMNRITAVTMEISVPYLCWPGRIQLOAK---VENPEAGOVNINVTFTS 357
 Db 301 DWVKAATR-----PYVLPGRQAVRCVPLPLTTQPPNIREKRGEGTLTPTMLMG 355
 QY 358 SQT 360
 Db 356 TOT 358
 RESULT 7
 Q8RGK2 PRELIMINARY; PRT; 1881 AA.
 AC Q8RGK2;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Hemolysin.
 GN FN0291.
 OS Fusobacterium nucleatum (subsp. nucleatum).
 OC Bacteria; Fusobacteriia; Fusobacteriales; Fusobacteriaceae;
 OC Fusobacterium
 OX NCBI_TaxID=76856;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 25586;
 MEDLINE-21886394; PubMed-11889109;
 RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
 Bhatnagar A., Bhatnagar A., Gardner W., Grechkin G., Zhu L.,
 Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
 Larsen N., D'Souza M., Malinas T., Pusch G., Haselkorn R.,
 Fongstein M., Kyripides N., Overbeek R.;
 RT "Genome sequence and analysis of the oral bacterium Fusobacterium
 nucleatum strain ATCC 25586."
 RT nucleatum strain ATCC 25586.
 RT J. Bacteriol. 184:2005-2018(2002).
 DR EMBL; AE010541; AAL94497.1; -.
 DR InterPro; IPR001452; SH3.
 DR InterPro; IPR000508; SigPase.
 DR Pfam; PF00018; SH3; 1.
 DR PROSITE; PS00761; SPASE_1_3; 1.
 KM Complete proteome.
 SQ SEQUENCE 1881 AA; 204375 MW; D635156A4EFA877 CRC64;
 Query Match 6.7%; Score 127; DB 16; Length 1881;
 Best Local Similarity 23.1%; Pred. No. 0.78;
 Matches 86; Conservative 45; Mismatches 120; Indels 122; Gaps 19;
 QY 1 MNKILFTLFFSVLETFV-----SADKIPGDESITNIFGPRDRNRSSEKHN- 49
 Db 5 LKKLIALFMLFHIISLADGIVDNSAKNLQVDRKANGVPLVNIAP-DNNGTS--HNW 61
 QY 50 -----ILNNHITAYSESHTLYDMNTFLCLSHNTLNGACPTSENPSSSVSG-- 96

Db 62 YKDYNDVGRGALNN-----SKDLNLSQAGLI 89
 QY 97 ENNTITQFEKRSLLKREL-----QIKYKQLLEKSVN-----PSGLTINSAHF--NKN 144
 Db 90 YGNPNLQNSKEASTIINEGVNKSRIEYQIYGRAYILANPGIYINAGFINGN 149
 QY 145 KNAAGSALYIYIP-----AGELKNLPFGGIMPATLKYRRRYSEYGYTITITIK 197
 Db 150 VFTTGRGNLNLNPEKGMIEING--KGLDLRINNAELIARAEISAPYIGGEVNNKLG 207
 QY 198 LTKGNIQIWLPOFSDARVDLNLPTGGGTYIGRNSVDMCFYDGTSTNSSLEIRFO-- 255
 Db 208 NQGSN---KPEYALDAR-----ALGSIYAGRINI-----IYVEDGVGYQAP 248
 QY 256 -----DNNPKSDGKEYLR-----KINDTKEIATYLSLLAGKSLPTNGTSLNI 301
 Db 249 MYATKGDVYISSKGYKYLKDTQAKDKIKISSTETELG--SKLAEINATIKSGKTSN-- 303
 QY 302 DASLETMMNRIT 314
 Db 304 -SGQIRAN--NNT 314
 RESULT 8
 Q8VU50 PRELIMINARY; PRT; 700 AA.
 ID Q8VU50
 AC Q8VU50;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Putative polymorphic membrane protein (Fragment).
 OS Chlamydia psittaci (Chlamydia phila psittaci).
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia phila.
 OX NCBI_TaxID=83554;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-POS;
 RA Laroucau K., Souriau A., Rodolakis A.;
 RT "Isolation of a new pmp sequence and evidence of pmp polymorphism in
 serotype-1 Chlamydia psittaci strains."
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF243415; AAL36959.1;
 DR InterPro; IPR003368; Chlamydia_pmp.
 DR Pfam; PF02415; DUF145; 1.
 DR TIGRPFAM; TIGR01376; POMP_repeat; 3.
 FT NON_TER 700
 SQ SEQUENCE 700 AA; 74200 MW; 2C38E504AF48D1D1 CRC64;
 Query Match 6.1%; Score 115.5; DB 2; Length 700;
 Best Local Similarity 18.3%; Pred. No. 1.7;
 Matches 73; Conservative 63; Mismatches 124; Indels 139; Gaps 18;
 QY 42 NESSPEHNLNNHITR-----YSESHTLYDMNTFLCLS--SHNTLNGAC----- 83
 Db 27 NETLSSDSYNGNVSDEFYKETTSGAIYTCBGNVCISYAGKDSPLNCSCTETENIS 86
 QY 84 -----PTSENPSSSVSGETNITLQFTEKRSLLKRELQIKGYKQLLEKSVN 129
 Db 87 FIGNGYTLCEFDNITTASNPALINSGD-----QKTLNVGGS--LESCAH 130
 QY 130 CPGLT-----LNSAHF--NCKM-----NAASGASLYIYIP-----A 159
 Db 131 CPPTGYGAIQTKGVSTFGNNKLFIDNNCSTGEGAIKATGSAELKLEGNVYVFS 190
 QY 160 GELKNLPFGGIMPATLKYRRRYSEYGYTITITIKLTDGNGNQIWLPOFGRSDARVDL 219
 Db 191 GNSSQKGGAIY--TKLITADGPTLFSNNNSVAS--SPKGG-AICLDITSSSCSLTA 244
 QY 220 NE-----RPTGGGTYIGRNSVDM-----CFYDGTSTNSSLE 251
 Db 245 NIGDITFGNKKYIKINGSSSYVKKRNAIDLGSCKFTKLNAKGFIFPHDPIANNGGSTE 304

```

QY 252 IRFODNNPSPD-----GK--FYLRKINDTKELAYL-----SLLA 286
      | | | | | | | | | | | | | | | | | | | | | | |
Db 305 IEL- NKTESDPTTYGKIYFSEGEKLSDEKTYPDMLKSYFKQPLKIGSLVJLKDGYTL 362
      | | | | | | | | | | | | | | | | | | | | | | |
QY 287 GKSLTPNCTSLNIDPALSLETNNKRIATVAMPETSLVEY 325
      | | | | | | | | | | | | | | | | | | | | | | |
Db 363 AKRIITQKSTVMDLGTTLQTPSSGETITLLFNLDINI 401
      | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 9

ID	PRELIMINARY;	PRT;	847 AA.
AC	P71132;		
DT	01-FEB-1997 (TrEMBLrel. 02, Created)		
DT	01-FEB-1997 (TrEMBLrel. 02, Last sequence update)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)		
DE	POMP91A.		
OS	Chlamydomophila abortus.		
OC	Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydomophila.		
OX	NCBI_TaxID=83555;		
PN	[1]		
PN	SEQUENCE FROM N.A.		
CC	STRAIN-S26/3;		
CC	MEDLINE=96406378; PubMed=8810511;		
RA	Longbottom D., Russell M., Jones G.E., Lainson F.A., Herring A.J.;		
RT	"Identification of a multigene family coding for the 90 kDa proteins		
RL	of the ovine abortion subtype of Chlamydia psittaci.";		
RL	FEWS Microbiol. Lett. 142:277-281(1996).		
RN	[2]		
RN	SEQUENCE FROM N.A.		
RP	STRAIN-S26/3;		
RC	MEDLINE=98187897; PubMed=9529048;		
RA	Longbottom D., Russell M., Dunbar S.M., Jones G.E., Herring A.J.;		
RT	"Molecular cloning and characterization of the genes coding for the		
RT	highly immunogenic cluster of 90 kilodalton envelope proteins from the		
RT	Chlamydia psittaci subtype that causes abortion in sheep."		
RL	Infect. Immun. 66:1317-1324(1998).		
DR	EMBL; U65942; AAC15921.1; -		
DR	InterPro; IPR006315; Autotransport.		
DR	InterPro; IPR003368; Chlamydia_PMP.		
DR	Pfam; PF02415; D0P145. 1.		
DR	TIGRPFAM; TIGR01376; POMP_Trepact; 3.		
DR	TIGRPFAM; TIGR01376; autotrans_hall; 1.		
QO	SEQUENCE 847 AA; 90694 MW; 754C958BE7P71179E CRC64;		

Query Match	6.1%;	Score 115.5;	DB 2;	Length 847;
Best Local Similarity	21.6%;	Pred. No. 2.1;		
Matches 94;	Conservative 58;	Mismatches 153;	Indels 131;	Gaps 21

Db 9 LISSSLIVANSYSEE--PDOKLTLSAHSYNGNTNSEP-----FNPBLTSSNGTIYVCTG 62
 QY 70 FCLLS-----SHNTLGNACPT-----SENSSSSVSEFNTITL 102
 Db 63 NCIIVAGLDGSGLSSSCFPTDAGNLSPFLGNYTLCPDNTITQSSHHGALISVGS--FNKT- 120
 QY 103 QTEFRSLIKRLQIKGYKQKLLFKSVNCPSLTLSA-----HFNCNK 145
 Db 121 -----LDLSGFS--LFSCAYCPGATGYGAKAVAGNTTIKDNSSLVFHKKCST 166
 QY 146 NMAAGSILYLYIPAGELK-----MLPF-----GGIMDATLKLRYKRRYSSETGYTYT 192
 Db 167 GEGGAIQCKASSSEAEKLEIENNQNLVFEANSSSSSGGAIYAD--KLTVISGGLPTFSNSV 225
 QY 193 NIT-----IKLTLDGNIQIWLPOKRSAPARYDLN--LRPGGGGYTIGRNVDM----- 237
 Db 226 SASPSRGKALICIKDSGG--ECSLTADLGDITDQKIIKIKTNGSPFTVARNISIDLGSGKFT 264
 QY 238 -----CYDGS--TNSSSELEIRQDNNPXSQK--FYLRKIMDTKETALYTL-- 281
 Db 285 KLNAGEGRIEFPYDITGGGDELIANNQDIPVDT--GIYFSGERLSDDEKKVKAANLKSD 343

```

QY      282  -----SILLG-----KSLTPNGSLINIAAASLET--NNNRIRIYAAMPETISV 323
          || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      344  EKQPLKIGSGSLIKDGVLTETKSTFTQTBGATVAMDCTLTQITSSGGERTITLINIDINV 403
          || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      324  PVLCPGRLQLDARKE 359
          || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      404  ASLGGGVAADPAARKE 419
          || | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 10

IT	08Q07H7;	PRELIMINARY;	PRT;	865 AA.
AC	08Q07H7;			
DT	01-JUN-2002 (TrEMBLrel. 21, Created)			
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)			
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)			
DE	Envelope glycoprotein (Fragment).			
EN	ENV.			
OS	Human immunodeficiency virus 1.			
OC	Viruses; Retroid viruses; Retroviridae; Lentivirus.			
OX	NCBI_TaxID=11676;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=99CMA304.			
RX	MEDLINE=21849375; PubMed=11860674;			
RA	Yamaguchi J., Vallari A.S., Swanson P., Bodelle P., Kaput L.,			
RA	Manosp C., Zekeng L., Guttler L.G., Devare S.G., Brennan C.A.;			
RT	"Evaluation of HIV Type 1 Group O Isolates: Identification of Five			
RT	Phylogenetic Clusters.";			
RL	AIDS Res. Hum. Retroviruses 18:269-282(2002).			
RL	EMBL; AF383244; AAL98866.1; -			
DR	InterPro: IPR000328; ENV_GP41.			
DR	InterPro: IPR000777; GP120.			
DR	Pfam: PF00516; GP120; 1.			
DR	Pfam: PF00517; GP41; 1.			
KW	AIDS; Coat protein; Glycoprotein; Polypeptide; Transmembrane.			
FT	NON_TER			
FT	SEQUENCE 865 AA; 97118 MW; 996B83862AA51ACA CRC64;			

Query Match	6.0%	Score 114	DB 15	Length 865
Best Local Similarity	21.5%	Pred. No. 2.9		
Matches 93	Conservative 56	Mismatches 150	Indels 134	Gaps 22

```

QY      21 VSAIDKIPGDEISTINIFGPDRNESSPKNNILNNH-----ITAYSESHYDNRNTEFL 71
Db      75 VPDPIPRPYEYPLAHNV-----IDKRNIMENITVKGMOGDIIDLMQESLKPVCNMFEL 125
QY      72 CUSNHT-LNGACPTSENBSSSVS---GETNITIDPTEKRSILKRELOIKGYKOLLEK 126
Db      126 CVOOMNTSLN---DFKSNTPSPENIMAKCEFNFTVAVKKDKKEKQALLFYVSDLMKIDDK 182
QY      127 SVNCPSEGLNSAHFNCKNNAASGA-----SLYLYIPAG---ELKNIPGCG--- 169
Db      183 NNTNTNTMTLT---NCNSTTISQACPRVSEEPPIHNCAPAGAIKCNNEENGITC 238
QY      170 -----IMDATL---KLRV-KRRYSERYG--TYTINITYIKLTOK--G 202
Db      239 KNTVVTCHGAKPTVSTQILNGFLNSGKIRIMTKNISDNVNITYLNTSLKATICKRPG 298
QY      203 NIOIWLPOKSPARVDLNRPTGGSTYGRNSVDMCTGYGTSTWSLSLETRFODNNPKSD 262
Db      299 NNTVOOVSIGPMWMSDLR---GTERNTSRVAYCEY---NSTWME-----RFLKOT 344
QY      263 GKPYLKRKIDDKREIAYTLISLLAG-----KSL---TPYNGTSL 298
Db      345 AERFLELVN-NKKKADMTFSSNGGDEPVALHNCHEEPYCTSSLEFNYTSSCMGSTC 403
QY      299 NIADAASLETNNKRI-----FAVTMEISVPVLCWPGRLQIDAKVENPEA 343
Db      404 NTKKSNN-STNTRIPCRLRQVYKSMIGGSGGLVAPPRLRGNTLWMSNITGLMLOMP-- 460
QY      344 GQYMGNIINTFTF 356

```

DB 461 ---WNHSNATFRP 470

RESULT 11

081521 PRELIMINARY; PRT; 2646 AA.

AC 081521; 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Erythrocyte membrane protein 1 (PFEMP1).
GN PEL1950W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.

RC MEDLINE-2225705; PubMed-12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Paul A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pereira M.A., Allen J., Selengut J., Haft D., Mather M.W., Valdivia A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum";
RL Nature 419:498-511(2002).
DR EMBL: AE014850; AA036476.1; -
SQ SEQUENCE 2646 AA; 300284 MW; C91D257F3DEF8717 CRC64;

Query Match Best Local Similarity 24.8%; Score 113; DB 5; Length 2646;
Matches 77; Conservative 37; Mismatches 128; Indels 68; Gaps 15;

QY 28 GDESTINIFGPRDRNESSPKHNLNHTAYSESHLYDRMTFLCISHTNLGACPTS- 86
DB 695 GGS--TAIGIKITLKKKKKKKKAN-----ISEKILDELFLNHELDKATCKCPKRF 747
QY 87 ENPSSSVSGENITITQTEKRSLLKRELQIKGYKQLLFKSVNCPGGLTNSHFNCKNN 146
DB 748 KNPCCSDTSGDSN--KQYAVANTVAOIIQGRKQKL-----HGSGSN 789
QY 147 AASGASLYLYIPAGELKNLPFGIMDATLKRKRYSEYGTITITIKLTKGNI-- 204
DB 790 ALKGNQNAKINNGRPN-----PLTACQITKHSNGK---DSNNPCNKKGNRLK 838
QY 205 --QIWLPOKSD-ARYDL-----NLRPTGGGYIGRNSVDMCFYDG--YSTNS 247
DB 839 IQQVW--SIKNDTSYDYVMPRRQHMCTSNLEKLYASVIGSNVNDKFLVEVLHAKS 896
QY 248 SSELFRQNNKPSCKKFLKIKINDT-KEIAYT---LSLLAGKSLPTPTNGSLINIA 303
DB 897 EAEFLKKKYNKQNGKGRKRDQATTCRAIRYSPADIGDIIRGKDLMDNN-----DA 950
QY 304 ASLETMMNRI 313
DB 951 KSLQTNLRAI 960

RESULT 12

08XNM2 PRELIMINARY; PRT; 743 AA.

AC 08XNM2; 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein CPE0220.
GN CPE0220.
OS Clostridium perfringens.

OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX PubMed-11792842;
RA Shimizu T., Ohnami K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kohara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL: AP003185; BAB79926.1; -
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR006635; NEA_transpt.
DR SMART: SM00725; NEAT. 4.
DR TIGRfams: TIGR01167; LPXNG_anchor. 1.
KW Hypothetical protein. Complete proteome.
SQ SEQUENCE 743 AA; 81149 MW; A7C451E9C5A5E545 CRC64;

Query Match Best Local Similarity 18.8%; Score 111.5; DB 16; Length 743;
Matches 64; Conservative 66; Mismatches 115; Indels 95; Gaps 13;

QY 7 ITTLFFSSVLFPPFANSAKIPGDESTITIFGPRDRNES---SPKHN-LNNHTAYSES 61
DB 88 LMTVFNSSLYGFNMNIEVSAGGEALKEIKENKDDSIPEVSPDKVKYIGLFTMMGRK 147
QY 62 HTLYDGMTFLCISHTLN--GACPTSENPSSSV-----SG-----ETNITLQ 103
DB 148 VELF-----LYNDMNTVLLDEAPLNNAKDISYTGCAIDLSGIVGTCKEDSNLKYE 201
QY 104 FTEKRSLLK-RELQIKGYKQLLFKSVNCPGGLTNSHFNCKNNKMAASGASLYLYIPAGEL 162
DB 202 ISGDTSFVADGKAIEIGVPIYTK-----VTDSSGQGD----- 234
QY 163 KNPFGIMDATLKRKRYSEYGTITITIKLTKGNIQ---ITLPOKSDARD 218
DB 235 -----EKTVMYVNNKKTTLGDSLYLKNTVYVGGNNETGNSARKVLYSDSRID 285
QY 219 LNLPTGGGYIGRNSVDMCFYDGYSTNSSLEIRQDNPNPKSDGKFLRKINDPTKEIA 278
DB 286 IS-----NKNNTVITLF-----NSLTAFLKNNVNVYDGEKVAEYKNNKRTIK 329
QY 279 YTL-----SLLAGKSLPTPTNGSLINIDAALEFN 309
DB 330 FNILPDLNSDIYVTLVSMGKKEVSFKTLNVTYAKLEDN 369

RESULT 13

08E9G6 PRELIMINARY; PRT; 2768 AA.

AC 08E9G6; 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE RTX toxin, putative.
GN SO4317.

OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=MR-1;
RA MEDLINE-22297686; PubMed-12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsaplin A., Scott J., Beaman M., Brinkac L., Daugherty S.,
RA Deboy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Imprimi M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouli H., Gill J., Uitterback T.R., McDonald L.A.,

Db	419	TLNSN--TGHC--TYLENAGTDNNLSNVASSNGVGIYL-VNSNNTALENNIATGNS	471
QY	248	SSLERFQDNNPKSDGKFLRKIN-DPTKEIAYTIS--LLAGKSLPTNGTSLNIADA	303
Db	472	KGIYVMTSNGNTISENEYNNNDVDVNSNGIMISLSNNKVSCKAYNNPYGISLNSSTN	531
QY	304	ASLETN	309
Db	532	NNISSN	537

Search completed: August 19, 2003, 09:48:13
Job time : 105 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 2003, 09:45:04 ; Search time 30 seconds
(without alignments)
509.141 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1886

Sequence: 1 MNKILFLFTLFSSVLETF...EAGQYMGNIWTFPSQTL 361

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 4231058 residues

328717

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cg2_6/ptodata/1/1aa/5A.COMB.pep.*
2: /cg2_6/ptodata/1/1aa/5B.COMB.pep.*
3: /cg2_6/ptodata/1/1aa/6A.COMB.pep.*
4: /cg2_6/ptodata/1/1aa/6B.COMB.pep.*
5: /cg2_6/ptodata/1/1aa/PTUS.COMB.pep.*
6: /cg2_6/ptodata/1/1aa/backfile1.pep.*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1536.5	81.5	360	2	US-08-483-101-17
2	988.5	52.4	363	2	US-08-483-101-16
3	927.5	49.2	364	2	US-08-483-101-5
4	116.5	6.2	1600	2	US-08-617-697-10
5	114	6.0	650	3	US-08-362-525-2
6	111.5	5.9	1005	4	US-09-206-942-41
7	111.5	5.9	1011	4	US-09-206-942-39
8	111.5	5.9	1011	4	US-08-728-470-10
9	111.5	5.9	1529	2	US-08-719-641-10
10	108	5.7	524	4	US-09-242-913B-15
11	107	5.7	1095	4	US-09-206-942-45
12	107	5.7	1101	4	US-09-206-942-43
13	106	5.6	671	2	US-08-737-716-13
14	101	5.4	915	4	US-09-206-942-35
15	101	5.4	1222	4	US-09-206-942-37
16	101	5.4	1228	4	US-09-206-942-34
17	101	5.4	1338	2	US-08-728-470-9
18	101	5.4	1338	3	US-08-719-641-9
19	101	5.4	1589	2	US-08-617-697-9
20	100.5	5.3	1004	4	US-09-206-942-57
21	100.5	5.3	1010	4	US-09-206-942-55
22	99.5	5.3	969	4	US-09-206-942-32
23	99.5	5.3	975	4	US-09-206-942-30
24	99	5.2	992	4	US-09-206-942-61
25	99	5.2	998	4	US-09-206-942-59
26	99	5.2	1848	3	US-08-296-791-6
27	99	5.2	1848	5	PCT-US95-10661A-6

28	97	5.1	977	4	US-09-206-942-53	Sequence 53, App1
29	97	5.1	983	4	US-09-206-942-51	Sequence 51, App1
30	94.5	5.0	1912	4	US-08-409-895-4	Sequence 4, App1
31	94.5	5.0	1912	3	US-08-685-467-4	Sequence 4, App1
32	94.5	5.0	2353	3	US-09-377-155-33	Sequence 33, App1
33	94.5	5.0	2353	3	US-08-913-942-4	Sequence 4, App1
34	94.5	5.0	2353	4	US-09-669-974-33	Sequence 33, App1
35	94.5	5.0	2354	4	US-09-268-347-47	Sequence 47, App1
36	94.5	5.0	2411	4	US-09-268-347-36	Sequence 36, App1
37	93.5	5.0	715	4	US-09-462-917A-134	Sequence 134, App
38	93.5	5.0	901	4	US-09-134-001C-5351	Sequence 5351, Ap
39	93	4.9	1167	2	US-08-485-568A-6	Sequence 6, App1
40	93	4.9	1167	2	US-08-590-554A-6	Sequence 6, App1
41	93	4.9	1167	2	US-09-184-523-6	Sequence 6, App1
42	92.5	4.9	990	4	US-09-252-991A-32469	Sequence 32469, A
43	92.5	4.9	1073	4	US-09-206-942-49	Sequence 49, App1
44	92.5	4.9	1079	4	US-09-206-942-47	Sequence 47, App1
45	91.5	4.9	869	2	US-08-483-101-15	Sequence 15, App1

ALIGNMENTS

RESULT 1
US-08-483-101-17
Sequence 17, Application US/08483101
Patent No. 5932715
GENERAL INFORMATION:
APPLICANT: Scott, June R.
APPLICANT: Froehlich, Barbara
TITLE OF INVENTION: CS2 Proteins and Coding Sequences
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee and Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,101
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 35878
REFERENCE/DOCKET NUMBER: 6-95
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 360 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-483-101-17
Query Match 81.5%; Score 1536.5; DB 2; Length 360;
Best Local Similarity 80.1%; Pred. No. 1e-149;
Matches 289; Conservative 33; Mismatches 38; Indels 1; Gaps 1;
QY 1 MNKILFLFTLFSSVLETF...EAGQYMGNIWTFPSQTL 361
Db 1 MNKILFLFTLFSSVLETF...EAGQYMGNIWTFPSQTL 361


```

OY 1 MNKLEIFLTFESSVLETFEAVASADKIPGDEITNIFGPRDRN--ESSPKNNIINHTAY 58
DB 1 LKRVLEIVLSMFLCSQVYGOASHHTNVEAGSINKTESIGPLDRSMAASYPAHYLFHEHVAGY 60
OY 59 SESHTLYDMRTFLCLSHNTLNGACPTSENSSSVSGEINTLTOTTERSLIKRELQIK 118
DB 61 NNDHSLFDMRTFLCLSHNTLNGACPTSENSSSVSGEINTLTOTTERSLIKRELQIK 118
OY 119 GKOLLEFKSVNC---PSGLTNSAHFNCKNAAGASLYLYIPAGELKMLPFGGIWDATL 175
DB 119 GKRFLEYESDRCHYVDKMNLSHNVKCVSGFTREGVDFLLYIPOGEIDILLGCIWEATL 178
OY 176 KLRVRKRYSETGYTYTINITIKLTDKNGIOIMLPKPSDARDVNLNRPFGGYYIGRNZY 235
DB 179 ELRVKRYHYNMGTIVNITVDLTDKNGIOIMLPKPSDARDVNLNRPFGGYYIGRNZY 238
OY 236 DMCFYDGYSTNSSLERFQDNPNKSGKRYLKRINDTKELAYTLSSLAGSLPTNG 295
DB 239 EMCLYDGYSTHSQSIEMRFQDQSGTGNNEYNLIKTGEPLKLPYKLSLLGGREFFPNNG 298
OY 296 TSLNTADASLETNMRITAVTPEISVYVLCMPGRLQDAKAVENPEAGQYMGNIWVTT 355
DB 299 EAFITNDYSSLEFINMRKISVSLPQISIVLCMPANLTPMSLEINPEAGEYSGILWVET 358
OY 356 PSSQT 361
DB 359 PSSSL 364

```

RESULT 4
US-08-617-697-10
Sequence 10, Application US/08617697
Patent No. 5977336

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J
TITLE OF INVENTION: High Molecular Weight Surface Proteins
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/617, 697
FILING DATE: 01-APR-1996

CLASSIFICATION:

CLASSIFICATION: 424

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US 08/302, 832
FILING DATE: 05-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Bekhtrester, Jerry W
REGISTRATION NUMBER: 22,651

REFERENCE/DOCKET NUMBER:

REFERENCE/DOCKET NUMBER: 1038-557

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813

INFORMATION FOR SEQ. ID NO. 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 1600 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
US-08-617-697-10

Query Match 6.2%; Score 116.5; DB 2; Length 1600;
Best Local Similarity 21.6%; Pred. No. 0.02;
Matches 74; Conservative 60; Mismatches 130; Indels 79; Gaps 19;

```

OY 48 HNLIINHTAYSESHLYIRMTFLCLSHNTLNGACPTSENSSSVSGEINTLT 102
DB 816 YNEYSKH--AINSHNL-----TILGVNVTIGG-----ENSSSITGININTNKNVTL 862
OY 103 OF-----TEKRSILIKRELQIKGYKOLLEFKSVNCPGSLT-----NSAHFNCN--- 144
DB 863 QADTSNNTGLKRTITLGNISVEGMLSTGANANIVGMLSTAEDESTFGEASDNINNTG 922
OY 145 ---KNAAGASLYLYIPAGELKMLPFGGIWDATLKLVRKRYSETGYT-YTINITIKLND 200
DB 923 TETNNGTANINIKGVYKLDINNK--GG-----LNTTNASGTQKTIINGNIT--NE 970
OY 201 KGNIOIMLPKPSDARDVNLNRPFGGYYIGRNZYDMCFYDGYSTNSSLERFQ 255
DB 971 KCDLNT--KNIRADAEIOLGNSISQEGMLTSSDKVNT-----TNOITKAGVEGR 1021
OY 256 -DNNPKSDGKFLYLRKINDTKELAYTLSSLAG---KSLPTNGTSLNTADASLETNWN 311
DB 1022 SDSESEANLNTIQ-----TKELKLAGDLNIGFKNKAEITAKNGSDTLTGMSGGNADAK 1076
OY 312 RITAVTPEISVYVLCMPG-RLQDAKAVENPEAGQYMGNIWVTT 353
DB 1077 K---VTFDKVKDSKISTDGHNVTLNSEVTSNGSSNAGNDNST 1116

```

RESULT 5
US-08-362-525-2
Sequence 2, Application US/08362525
Patent No. 6027910

GENERAL INFORMATION:

APPLICANT: KLIIS, FRANCISCUS M.
APPLICANT: SCHREIDER, MAARTEN P.
APPLICANT: TOSCHKA, HOLLER Y.
TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE
TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:

ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/362, 525
FILING DATE: 04-JAN-1995

CLASSIFICATION:

CLASSIFICATION: 435

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: EP 92202080.5
FILING DATE: 08-JUL-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: EP 92203899.7
FILING DATE: 14-DEC-1992

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813

INFORMATION FOR SEQ. ID NO. 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 1600 amino acids
TYPE: amino acid

```

:
:      REGISTRATION NUMBER: 16,773
:      REFERENCE/DOCKET NUMBER: 213289/77020(V)
:
: TELECOMMUNICATION INFORMATION:
:
:      TELEPHONE: (202) 861-3000
:      TELEFAX: (202) 822-0944
:      TELEX: 6714627 CUSH
:
:      INFORMATION FOR SEQ ID NO: 2:
:
:      SEQUENCE CHARACTERISTICS:
:
:          LENGTH: 650 amino acids
:          TYPE: amino acid
:          TOPOLOGY: linear
:
:      MOLECULE TYPE: protein
:
: US-08-362-525-2

```

Query Match	6.0%;	Score 114;	DB 3;	Length 650;
Best Local Similarity	20.9%;	Pred. No. 0.0085;		
Matches 98;	Conservative 52;	Mismatches 147;	Indels 172;	Gaps 23

```

0Y 44 SSP---KHNILNNHITA-----YSESHLYDRMTFLC---LSSHNTLN 80
Db 68 SMFHVRIKLNNSSCATATSLADGETAEKCYVSCQAALYENTFETCTCAQNDNLSYNTID 127
0Y 81 GACPTSENRSSSVSGETNITLQTFEKSILKRELIQINGYQOLLEKSYNC-PSGLTINSA 139
Db 128 GSTTFEFLNSDGGSSYEYL-----ENAKFFSGPMALIKLQNDMSDVNEDPAFTNVF 182
0Y 140 HFNCNKNASGASLYL-YIPAGELKNLPFGIMPATLKLHVKKRRSEYTYVINTIK 197
Db 183 HSGRSTGYGSPSEYHGMYCPRNGYF---LGS---TEKI-----DIVSSNNNDL- 225
0Y 198 LTRKGIQI-----WLPQFKDARVDLNRPTGGCITYGRN--SYDMCTDYG-- 242
Db 226 --DCSSQVYSSNDEPNMWFPOSYNDTNADV-----TFCGSLMTLDEKLYDEML 275
0Y 243 -----YSTNSSLE--IIPRODN- PKSDGF 265
Db 276 WVAALOSLPANNITDHALFQYTCLDITANTTYATQPSITRRETYVYQGRNLGIASKSS 335
0Y 266 YLRKINDDTKEI--AYTLS-----LLAGKSLPTPTNGSLINIDAA 304
Db 336 FISTTTTDLTSLNTSAVSGSISFVETGNRTTSEVISHVVTSTKSLPFAATSLTIQTS 395
0Y 305 SLETMNN-----RIAYVMPESIVYLCWPRGL 332
Db 396 IYSTDNITVGTDIHTTSEVISOVETISRETSIVAAVAPUSITGTGWTGM 444

```

```

RESULT 6
US-09-206-942-41
; Sequence 41, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; FILE REFERENCE: 1038-861 MIS:jb
; CURRENT APPLICATION NUMBER: US/09/206,942
; CURRENT FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: 09/167,568
; EARLIER FILING DATE: 1998-10-07
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 41
; LENGTH: 1005
; TYPE: PRF
; ORGANISM: Haemophilus influenzae
; US-09-206-942-41

```

Query Match	5.9%;	Score 111.5;	DB 4;	Length 1005;
Best Local Similarity	20.3%;	Pred. No. 0.031;		
Matches 68;	Conservative 65;	Mismatches 147;	Indels 55;	Gaps 14;

```

OY 60 ESTLTDRTMPTCLTSSHTNLNACPTSESSSPSSGCTNTLTQFTEKRS--LIRREJOIK 118
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 363 EKKAISTEHTNLTLTIGANTVLIGSESSSNIKGINININSKANVTLOAHAGTSHLDKE----418
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
OY 119 GYKOLLFFKSVNCPSCGTLT--NSAHFNCKMKNASGASLYLTPAGEL--KNLPFGGIW--DA 173
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 419 --RTLLIGANVSOGANNITIGSNAHIDGNLSIAESAKF-----OGKTNNNNLNITCTFINNG 471
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
OY 174 TLKLRVKRRYSEYGYTPTINTPTIKLTD-----KGNL-----OIWLPQFSKARDV 218
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 472 TADINIKOGVAVKOGDITNGNGLNNTTNASVNOQTIIINGNTNKKGDPLNDIKADANEIQ 531
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
OY 219 L--NLRPFGGTYIGANSYDMCFYDGYSTSSSLEIRPODNPXSD--GKFLYRKINDOTK 275
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 532 IGENISQKEGNLTISSDKINI-----TKRIEIKADTDQGSNDSGVSANMLTIYKT 582
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
OY 276 ELAYTSLSLLAG---KSLLPTMGTSLNITADAASLETNRNRTATYATMEISIVPLCWGR 332
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 583 ELTLTNTLNLISGFKNAEITAKDNSDLIIGKASSDNSNAKOITFDKVDXDSKIS--AGNHNV 640
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
OY 333 OLDARVENPEA-----GOYMGNIWVTFPPSQTL 361
   | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 641 TLNSKVEITNSDGSCTGNSGDDNNIGLTLISAKDYI 675

```

```

RESULT 7
US-09-206-942-39
Sequence 39, Application US/09206942
Patent No. 643269
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M.
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H.
TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
FILE REFERENCE: 1038-861 MIS:jb
CURRENT APPLICATION NUMBER: US/09/206,942
CURRENT FILING DATE: 1998-12-08
EARLIER APPLICATION NUMBER: 09/167,568
EARLIER FILING DATE: 1998-10-07
NUMBER OF SEQ ID NOS: 95
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 39
LENGTH: 1011
TYPE: PRF
ORGANISM: Haemophilus influenzae
US-09-206-942-39

```

[illegible]

QY 276 EIAVTLISLLAG---KSLPTNGTSLNIDAALETFNNRITAVTPEISVPLCWPGRL 332
DB 589 EITLDNINISGFNNAETAKNDSLLIGKASDSNMKQITFDKVKOSKIS--AGNHV 646
QY 333 OLDARVENPEA-----GOYMGINVTFTPPSSQTL 361
DB 647 TLNSKVEITSNDSGSGSDNNIGLITLSAKDVTY 681

RESULT 8

US-08-728-470-10
; Sequence 10, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: Of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-728-470-10

Query Match 5.9%; Score 111.5; DB 2; Length 1529;
Best Local Similarity 20.7%; Pred. No. 0.061;
Matches 72; Conservative 62; Mismatches 127; Indels 87; Gaps 18;

QY 48 HNILNNHTAYSESTLIDRMFTLCLSSNITNGACPSENSSS-----SYSGETNTTL 102
DB 744 YNEYSKH--AINSSHNL-----TLGGNVTLLG-----ENSSSITGNINITNANVTL 790
QY 103 QF-----TEKRLIKRELQIKGYKOLLFKSVNCPGSLTJNSAHFNCNKAASGASLY 154
DB 791 QADTSNMTGLKRLTLTGNTISVEGNLSITGNANITVGLST--AEDSTFKELAS----- 843
QY 155 LYIPAGELKNLPFGGIW--DATLKLVRKRRYSEYGYTV---INITTKLT----- 199

DB 844 -----DNLTNGTFTNGTANINIKOGVYKLOGDINNKGGLNITTNASGTQRTIING 895
QY 200 ----DKGNIQYLPQPKSARVDL--NLRPFGGTYIGNSVDMCFYDGYSTNSSLER 253
DB 896 NITNEKGDINT--KNIKDAEIOIGNISOKEGNTLTJISDKYNI-----TNQTTIAG 946
QY 254 FQ-----DNNPKSDGFYLRKINDPTKEIAYVTLISLLAG---KSLPTNGTSLNIDAA 306
DB 947 VEGGRSDSSEANALITLO-----TKELKAGDLNISGFNKRKREITAKKNSDLTIGNASG 1001
QY 307 ETNNNRITAVTPEISVPLCWPG--RLQDARVENPEAGQYGNINVT 353
DB 1002 NADAKK---VTEDKVKDSKISTDGHNVTLINSEVKTSSNGSSNAGDNDST 1046

RESULT 9

US-08-719-641-10
; Sequence 10, Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: Of No. 6218141-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-719-641-10

Query Match 5.9%; Score 111.5; DB 3; Length 1529;
Best Local Similarity 20.7%; Pred. No. 0.061;
Matches 72; Conservative 62; Mismatches 127; Indels 87; Gaps 18;

QY 48 HNILNNHTAYSESTLIDRMFTLCLSSNITNGACPSENSSS-----SYSGETNTTL 102
DB 744 YNEYSKH--AINSSHNL-----TLGGNVTLLG-----ENSSSITGNINITNANVTL 790

Db 489 GEVKSASGNVNTASGNTLNVSNITGQVTVTAN-----SGAITTEGST----- 533
QY 332 LQIDAKVENPEAGQYMNIN-----VFTEPSSQTL 361
Db 534 --INATGDNITTTGTGNGKVESSSGSVTLIATGQTL 570

RESULT 15

US-09-206-942-37
; Sequence 37, Application US/09206942
; Patent No. 6432669
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
; TITLE OF INVENTION: Molecular Weight Proteins
FILE REFERENCE: 1038-861 M1S:jb
CURRENT APPLICATION NUMBER: US/09/206,942
CURRENT FILING DATE: 1998-12-08
EARLIER APPLICATION NUMBER: 09/167,568
EARLIER FILING DATE: 1998-10-07
NUMBER OF SEQ ID NOS: 95
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 37
LENGTH: 1222
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-09-206-942-37

Query Match 5.4%; Score 101; DB 4; Length 1222;
Best Local Similarity 21.1%; Pred. No. 0.51;
Matches 84; Conservative 48; Mismatches 129; Indels 138; Gaps 19;

QY 21 VSADKIPGDESTINIFGR-----DRNESSPKH-----NILN-----NHITAYS 59
Db 559 ISSDVKVNTERTITKAGVNGDSNSNEATSANLTKTKELKLTNDLNSGENKAEITAKD 618
QY 60 ESH-TLYD-----RMTF-----LCISHN-TLNGACPTSENPSSSSVSGETNI 100
Db 619 NSNLIGNSDAGNDKAKVFSNVKDSKISASDHNVTLNSKVEISGDTSTEDGNNNT 678
QY 101 TLQTEKRSILKRELQIKYQOL--LFSVNCPSGLTLNSAHFNCKNNAASGASLYLYI 157
Db 679 GLTITAKNVTVNN--NITSKVTNITASENVTTKAGTTIN-----ATTGSVEVYA 726
158 PAGELKNLPFGIMDATLKLKVKRYSFTYTTITITIKLTKDNGNIQIWLPOFKSDARV 217
Db 727 KTGDIK---GGI-----ESNSGNVNI----- 744
QY 218 DLNLKPTGGTYI-----GRNSVDMCFYDGYSTNSSLEIRFQDNNPKSDGKFLRKIN 271
Db 745 -----TASGDTLNVSNITGQNVVAASGAVTTTKGSTINATGNNITTK--TGEIN 795
QY 272 DDTEKIAITLISLLAGKSILTPNGSLNIADAASLETMMNRITAVTMEPISVPLCWPER 331
Db 796 GEVKSASGNVNTASGNTLNVSNITGQVTVTAN-----SGAITTEGST----- 840
QY 332 LQIDAKVENPEAGQYMNIN-----VFTEPSSQTL 361
Db 841 --INATGDNITTTGTGNGKVESSSGSVTLIATGQTL 877

Search completed: August 19, 2003, 09:49:35
Job time : 32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 2003, 09:48:19 : Search time 57 Seconds
(without alignments)
829.687 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1886
Sequence: 1 MNKILFFLFSSVLFETFA.....EAGQVGNINVTFFSSQTL 361

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 492763 seqs, 131003257 residues

Total number of hits satisfying chosen parameters: 492763

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications-AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEM_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1886	100.0	361	US-09-839-894-10	Sequence 10, Appl
2	1862.5	98.8	361	US-09-839-894-28	Sequence 28, Appl
3	1542	81.8	359	US-09-839-894-29	Sequence 29, Appl
4	937	49.7	362	US-09-839-894-31	Sequence 31, Appl
5	916.5	48.6	364	US-09-839-894-30	Sequence 30, Appl
6	247.5	13.1	353	US-09-839-894-32	Sequence 32, Appl
7	119.5	6.3	770	US-09-303-232-2	Sequence 2, Appl
8	116.5	6.2	1600	US-10-092-880-10	Sequence 10, Appl
9	111.5	5.9	1005	US-10-193-764-41	Sequence 41, Appl
10	111.5	5.9	1011	US-10-193-764-39	Sequence 39, Appl
11	109	5.8	2383	US-09-912-020-302	Sequence 302, Appl
12	108	5.7	524	US-10-347-278-15	Sequence 15, Appl
13	101	5.4	915	US-10-193-764-35	Sequence 35, Appl
14	101	5.4	1222	US-10-193-764-37	Sequence 37, Appl
15	101	5.4	1228	US-10-193-764-34	Sequence 34, Appl

16	101	5.4	1599	US-10-092-880-9	Sequence 9, Appl
17	100.5	5.3	1004	US-10-193-764-53	Sequence 53, Appl
18	100.5	5.3	1010	US-10-193-764-51	Sequence 51, Appl
19	99.5	5.3	969	US-10-193-764-32	Sequence 32, Appl
20	99.5	5.3	975	US-10-193-764-30	Sequence 30, Appl
21	99	5.2	839	US-10-023-437-23	Sequence 23, Appl
22	99	5.2	992	US-10-193-764-57	Sequence 57, Appl
23	99	5.2	998	US-10-193-764-55	Sequence 55, Appl
24	99	5.2	1848	US-09-839-894-6	Sequence 6, Appl
25	99	5.2	1848	US-10-080-505-6	Sequence 6, Appl
26	97	5.1	977	US-10-193-764-49	Sequence 49, Appl
27	97	5.1	983	US-10-193-764-47	Sequence 47, Appl
28	95.5	5.1	825	US-10-101-464-922	Sequence 922, Appl
29	95	5.0	263	US-09-839-894-6	Sequence 6, Appl
30	94.5	5.0	2353	US-09-797-862-33	Sequence 33, Appl
31	92.5	4.9	1073	US-10-193-764-45	Sequence 45, Appl
32	92.5	4.9	1079	US-10-193-764-43	Sequence 43, Appl
33	91.5	4.9	1143	US-09-924-154-14	Sequence 14, Appl
34	91.5	4.9	1236	US-09-769-787-109	Sequence 109, Appl
35	91	4.8	1395	US-10-080-505-7	Sequence 7, Appl
36	90.5	4.8	564	US-09-944-160-12	Sequence 12, Appl
37	90.5	4.8	1019	US-09-776-191-64	Sequence 64, Appl
38	90.5	4.8	1019	US-10-157-031-267	Sequence 267, Appl
39	89.5	4.7	811	US-10-011-588-29	Sequence 29, Appl
40	89	4.7	427	US-10-214-524-36	Sequence 36, Appl
41	89	4.7	465	US-09-957-485-6	Sequence 6, Appl
42	89	4.7	465	US-09-987-021-6	Sequence 6, Appl
43	88	4.7	921	US-09-117-447-6	Sequence 6, Appl
44	88	4.7	1220	US-10-193-764-28	Sequence 28, Appl
45	88	4.7	1226	US-10-193-764-26	Sequence 26, Appl

ALIGNMENTS

US-09-839-894-10	US-10-092-880-9
Sequence 10, Application us/09839894	
Patent No. US20020176868A1	
GENERAL INFORMATION:	
APPLICANT: Alboum, Zeev	
APPLICANT: Barry, Eileen M.	
APPLICANT: Levine, Myron M.	
APPLICANT: University of Maryland	
TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE	
FILE REFERENCE: USFMD.006A	
CURRENT APPLICATION NUMBER: US/09/839,894	
CURRENT FILING DATE: 2001-04-20	
PRIOR APPLICATION NUMBER: 60/198,626	
PRIOR FILING DATE: 2000-04-20	
NUMBER OF SEQ ID NOS: 40	
SOFTWARE: FastSeq for Windows Version 4.0	
SEQ ID NO 10	
LENGTH: 361	
TYPE: PRT	
ORGANISM: E. coli	
US-09-839-894-10	US-10-092-880-9
Query Match	100.0%; Score 1886; DB 10; Length 361;
Best Local Similarity	100.0%; Pred. No. 1.5e-181;
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 MNKILFFLFSSVLFETFAVSADKIPGDESTINIGPRDRNESSPKHNLNNHTAYSE 60
DB	1 MNKILFFLFSSVLFETFAVSADKIPGDESTINIGPRDRNESSPKHNLNNHTAYSE 60
OY	SHLYLRMTFLCSSHNTNGACPTSENPSSSVSEGTITLOFTKRSIRKRELOIKY 120
DB	SHLYLRMTFLCSSHNTNGACPTSENPSSSVSEGTITLOFTKRSIRKRELOIKY 120
OY	KOLFRVNCPSGLTLNSAHFNCNKNAASGASLYLTPAGELNLPFGGITWATKLRYK 180

```

Db      121 KOLLKSVNCPGSLTNSAHFNCKNNAASGASLYIYPAGELKNI.PFGGIMDATIKLRVK 180
Qy      181 RRYSETGYTYTINITIKITLDKGNIOIWLPOFKSDARVDLNLPTGGTYIGNSVDMCFY 240
        181 RRYSETGYTYTINITIKITLDKGNIOIWLPOFKSDARVDLNLPTGGTYIGNSVDMCFY 240
Db      241 DGYSTNSSSLERFDONNPKSDGKFKYLRKINDTEIAYTISLLAGSLPTNGTSLNI 300
Qy      241 DGYSTNSSSLERFDONNPKSDGKFKYLRKINDTEIAYTISLLAGSLPTNGTSLNI 300
        241 DGYSTNSSSLERFDONNPKSDGKFKYLRKINDTEIAYTISLLAGSLPTNGTSLNI 300
Db      301 ADAASLETNNMRTAVTAVTPEISVPLCWPGRLQDAKVENPAGQYMGNIINVTFTPSST 360
Qy      301 ADAASLETNNMRTAVTAVTPEISVPLCWPGRLQDAKVENPAGQYMGNIINVTFTPSST 360
        301 ADAASLETNNMRTAVTAVTPEISVPLCWPGRLQDAKVENPAGQYMGNIINVTFTPSST 360
Db      361 L 361
Qy      361 L 361
        361 L 361

```

Query 2

```

US-09-839-894-28
; Sequence 28, Application US/09839894
; Patent No. US20020176868A1
; GENERAL INFORMATION:
; APPLICANT: Alboum, Zeev
; APPLICANT: Barry, Eileen M.
; APPLICANT: Levine, Myron M.
; APPLICANT: University of Maryland
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
; FILE REFERENCE: USFMD 006A
; CURRENT APPLICATION NUMBER: US/09/839,894
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/198,626
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: EREC Protein Homology Sequence
US-09-839-894-28

```

Query Match 98.8%; Score 1862.5; DB 10; Length 361;

Best Local Similarity 99.4%; Pred. No. 3.6e-179;

Matches 359; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```

Qy      1 MNKILFIFTLFFSSVLTFFAVSADKIPGDESITNIFGPRDRNESSPKHNIINNHITAYSE 60
        1 MNKILFIFTLFFSSVLTFFAVSADKIPGDESITNIFGPRDRNESSPKHNIINNHITAYSE 61
Db      2 MNKILFIFTLFFSSVLTFFAVSADKIPGDESITNIFGPRDRNESSPKHNIINNHITAYSE 61
Qy      61 SHTIDRMTFLCLSHNTLNACPTSENPSSSVSGENITNLOFTEKSLIKRELOIKGY 120
        61 SHTIDRMTFLCLSHNTLNACPTSENPSSSVSGENITNLOFTEKSLIKRELOIKGY 120
Db      62 SHTIDRMTFLCLSHNTLNACPTSENPSSSVSGENITNLOFTEKSLIKRELOIKGY 121
Qy      121 KOLLKSVNCPGSLTNSAHFNCKNNAASGASLYIYPAGELKNI.PFGGIMDATIKLRVK 180
        121 KOLLKSVNCPGSLTNSAHFNCKNNAASGASLYIYPAGELKNI.PFGGIMDATIKLRVK 180
Db      122 KOLLKSVNCPGSLTNSAHFNCKNNAASGASLYIYPAGELKNI.PFGGIMDATIKLRVK 181
Qy      181 RRYSETGYTYTINITIKITLDKGNIOIWLPOFKSDARVDLNLPTGGTYIGNSVDMCFY 240
        181 RRYSETGYTYTINITIKITLDKGNIOIWLPOFKSDARVDLNLPTGGTYIGNSVDMCFY 240
Db      182 RRYSETGYTYTINITIKITLDKGNIOIWLPOFKSDARVDLNLPTGGTYIGNSVDMCFY 241
Qy      241 DGYSTNSSSLERFDONNPKSDGKFKYLRKINDTEIAYTISLLAGSLPTNGTSLNI 300
        241 DGYSTNSSSLERFDONNPKSDGKFKYLRKINDTEIAYTISLLAGSLPTNGTSLNI 300
Db      242 DGYSTNSSSLERFDONNPKSDGKFKYLRKINDTEIAYTISLLAGSLPTNGTSLNI 300
Qy      301 ADAASLETNNMRTAVTAVTPEISVPLCWPGRLQDAKVENPAGQYMGNIINVTFTPSST 360
        301 ADAASLETNNMRTAVTAVTPEISVPLCWPGRLQDAKVENPAGQYMGNIINVTFTPSST 360
Db      301 ADAASLETNNMRTAVTAVTPEISVPLCWPGRLQDAKVENPAGQYMGNIINVTFTPSST 360

```

```

Qy      361 L 361
Db      361 L 361

```

RESULT 3

```

US-09-839-894-29
; Sequence 29, Application US/09839894
; Patent No. US20020176868A1
; GENERAL INFORMATION:
; APPLICANT: Alboum, Zeev
; APPLICANT: Barry, Eileen M.
; APPLICANT: Levine, Myron M.
; APPLICANT: University of Maryland
; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
; FILE REFERENCE: USFMD 006A
; CURRENT APPLICATION NUMBER: US/09/839,894
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/198,626
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: EREC Protein Homology Sequence
US-09-839-894-29

```

Query Match 81.8%; Score 1542; DB 10; Length 359;

Best Local Similarity 80.9%; Pred. No. 7.3e-147;

Matches 292; Conservative 32; Mismatches 35; Indels 2; Gaps 2;

```

Qy      1 MNKILFIFTLFFSSVLTFFAVSADKIPGDESITNIFGPRDRNESSPKHNIINNHITAYSE 60
        1 MNKILFIFTLFFSSVLTFFAVSADKIPGDESITNIFGPRDRNESSPKHNIINNHITAYSE 60
Db      1 MNKILFIFTLFFSSVLTFFAVSADKIPGDESITNIFGPRDRNESSPKHNIINNHITAYSE 60
Qy      61 SHTIDRMTFLCLSHNTLNACPTSENPSSSVSGENITNLOFTEKSLIKRELOIKGY 120
        61 SHTIDRMTFLCLSHNTLNACPTSENPSSSVSGENITNLOFTEKSLIKRELOIKGY 120
Db      61 SHTIDRMTFLCLSHNTLNACPTSENPSSSVSGENITNLOFTEKSLIKRELOIKGY 120
Qy      121 KOLLKSVNCPGSLTNSAHFNCKNNAASGASLYIYPAGELKNI.PFGGIMDATIKLRVK 180
        121 KOLLKSVNCPGSLTNSAHFNCKNNAASGASLYIYPAGELKNI.PFGGIMDATIKLRVK 180
Db      121 KOLLKSVNCPGSLTNSAHFNCKNNAASGASLYIYPAGELKNI.PFGGIMDATIKLRVK 180
Qy      181 RRYSETGYTYTINITIKITLDKGNIOIWLPOFKSDARVDLNLPTGGTYIGNSVDMCFY 240
        181 RRYSETGYTYTINITIKITLDKGNIOIWLPOFKSDARVDLNLPTGGTYIGNSVDMCFY 240
Db      181 RRYSETGYTYTINITIKITLDKGNIOIWLPOFKSDARVDLNLPTGGTYIGNSVDMCFY 239
Qy      241 DGYSTNSSSLERFDONNPKSDGKFKYLRKINDTEIAYTISLLAGSLPTNGTSLNI 300
        241 DGYSTNSSSLERFDONNPKSDGKFKYLRKINDTEIAYTISLLAGSLPTNGTSLNI 300
Db      240 DGYSTNSSSLERFDONNPKSDGKFKYLRKINDTEIAYTISLLAGSLPTNGTSLNI 299
Qy      301 ADAASLETNNMRTAVTAVTPEISVPLCWPGRLQDAKVENPAGQYMGNIINVTFTPSST 360
        301 ADAASLETNNMRTAVTAVTPEISVPLCWPGRLQDAKVENPAGQYMGNIINVTFTPSST 360
Db      300 ADAASLETNNMRTAVTAVTPEISVPLCWPGRLQDAKVENPAGQYMGNIINVTFTPSST 358
Qy      361 L 361
        361 L 361
Db      359 L 359

```

RESULT 4

```

US-09-839-894-31
; Sequence 31, Application US/09839894
; Patent No. US20020176868A1
; GENERAL INFORMATION:
; APPLICANT: Alboum, Zeev
; APPLICANT: Barry, Eileen M.
; APPLICANT: Levine, Myron M.
; APPLICANT: University of Maryland

```

```

: TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
: FILE REFERENCE: CSA OPERON
: CURRENT APPLICATION NUMBER: US/09/839,894
: CURRENT FILING DATE: 2001-04-20
: PRIOR APPLICATION NUMBER: 60/198,626
: PRIOR FILING DATE: 2000-04-20
: NUMBER OF SEQ ID NOS: 40
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 31
: LENGTH: 362
: TYPE: PR
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: ETEC Protein Homology Sequence
: US-09-839-894-31

Query Match          49.7%; Score 937; DB 10; Length 362;
Best Local Similarity 53.1%; Pred. No. 7.5e-86;
Matches 197; Conservative 54; Mismatches 98; Indels 22; Gaps 12;

3 KILFIFLFFSVLFTFAVSADKIPGDE--STINIF-GPR-DRNESSPKHNLNHTAY 58
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
2 KKIIFL---SIIFSAVVSAGRPETTVGNLTKSFQAPRIDRSVQSPIYINIFNHVAGY 57
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
59 SESHLYDRMTFLCLSHNTLNGACPTSENPSSSVS-GFTNITLOTEKRSILKRELQI 117
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
58 SLHSISLYDRIVFLCTSSNRYNGACPTI--GTSQVGYGTTTITLOTEKRSILKRNINI 114
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
118 KGYKOLLKFSVNCPSG--LTLNSAHFNCNKA--ASGASLYLYIPAGELKNLPFGIWDAT 174
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
115 AGNKKPIWENOSCDPFSINIVLNSKMSGCAHGNANGTILNLYIPAGEINKLPFGIWEAT 174
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
175 LKLRKRREYSET---GTYTINTITLTKDGNQIQLPQKSDARVDLNLPRGGTYI 230
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
175 LILRLLS-RIGEVSTHGTNTVTLDTKGNQVLPQFHSNPRVDLNLPRIGNKYS 233
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
231 GRNSVDMCFYDGYSTNSSLEIRFODNNPKSDGKFLYLRKINDTKEIAYTLLSLLAGSL 290
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
234 GSNSLDMCFYDGYSTNDSNVYIKFODNPTNSSSYNLYKYG-GIEKLPYAVS-LIGKIF 291
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
291 TTPNGISLNIADAASLETNNRITAVTPEISVPLVLCWPGRLQIDAKVENPEAGQYNGNT 350
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
292 YPVNGOSFTINDSVLETNNRVTAVAMPENVPVLCWPARLLNADVNA PDAGQYSGQI 351
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
351 NVTFTPSQTL 361
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
352 YITFTPSVENTL 362
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 5
: US-09-839-894-30
: Sequence 30, Application US/09839894
: Patent No. US20020176868A1
: GENERAL INFORMATION:
: APPLICANT: Alboum, Zeev
: APPLICANT: Barry, Eileen M.
: APPLICANT: Levine, Myron M.
: APPLICANT: University of Maryland
: TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
: FILE REFERENCE: CSA OPERON
: CURRENT APPLICATION NUMBER: US/09/839,894
: CURRENT FILING DATE: 2001-04-20
: PRIOR APPLICATION NUMBER: 60/198,626
: PRIOR FILING DATE: 2000-04-20
: NUMBER OF SEQ ID NOS: 40
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 30
: LENGTH: 364
: TYPE: PR
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: ETEC Protein Homology Sequence
: US-09-839-894-30

```

```

: OTHER INFORMATION: ETEC Protein Homology Sequence
: US-09-839-894-30

Query Match          48.6%; Score 916.5; DB 10; Length 364;
Best Local Similarity 48.1%; Pred. No. 8.8e-84;
Matches 176; Conservative 63; Mismatches 120; Indels 7; Gaps 3;

1 MKRLIFLTLFFSVLFTFAVSADKIPGDESITRIFPRDRN--ESSPKHNLNHTAY 58
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1 MKRVIFLMSFLSQYVGQSMHNVGASINKFTSITPIDRSAASAPAHYIFHEBVAGY 60
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
59 SESHLYDRMTFLCLSHNTLNGACPTSENPSSSVS-GFTNITLOTEKRSILKRELQI 118
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
61 NKHSLEFDRMTFLCMSSTDASKACPIGESSKSS--QGFNLIKILFEKSLARKTILNK 118
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
119 GYKOLLKFSVNC--PSGLTNSAHFNCNKAASGASLYLYIPAGELKNLPFGIWDATL 175
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
119 GYKRFLESYDRCHIVYDKMNLNHTVACVGSFTFGVDFLYIPQGEIDGLTGQYWKATL 178
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
176 KLRVRRYSETYGTITNITIKLTKDGNQIQLPQKSDARVDLNLPRGGTYIGRNSV 235
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
179 ELRVKRRYDYNHGTYKVNITVLDKGNIQVWPQFHSNPRVDLNLPRIGNKYS 238
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
236 DMCFYDGYSTNSSLEIRFODNNPKSDGKFLYLRKINDTKEIAYTLLSLLAGSLPPTNG 295
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
239 EMCICYDSTHSGSIEHRFODDSOTGNNEINLYKTEGPLKLPKXSLLLGREFYPNG 298
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
296 TSLNIADAASLETNNRITAVTPEISVPLVLCWPGRLQIDAKVENPEAGQYNGNT 355
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
299 KAFINDTSSLFTIMNRKISVLSIPVLCWPARLTFMSELNPEAGEYSGILNVTFT 358
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
356 PSSQTL 361
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
359 PSSSSL 364
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 6
: US-09-839-894-32
: Sequence 32, Application US/09839894
: Patent No. US20020176868A1
: GENERAL INFORMATION:
: APPLICANT: Alboum, Zeev
: APPLICANT: Barry, Eileen M.
: APPLICANT: Levine, Myron M.
: APPLICANT: University of Maryland
: TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
: FILE REFERENCE: CSA OPERON
: CURRENT APPLICATION NUMBER: US/09/839,894
: CURRENT FILING DATE: 2001-04-20
: PRIOR APPLICATION NUMBER: 60/198,626
: PRIOR FILING DATE: 2000-04-20
: NUMBER OF SEQ ID NOS: 40
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 32
: LENGTH: 353
: TYPE: PR
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: ETEC Protein Homology Sequence
: US-09-839-894-32

Query Match          13.1%; Score 247.5; DB 10; Length 353;
Best Local Similarity 28.9%; Pred. No. 2.4e-16;
Matches 89; Conservative 42; Mismatches 126; Indels 51; Gaps 13;

77 NTLNG-ACPTSENPSSS-----VSGETNITLOTEKRSILKRELQIKGYKOLLFK 126
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
72 NVLGGWYCRSNRNRNENECETHLWYAFAYISIRLRFRRQJSHAETLL----- 120
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
127 SVNCPGSLTNSAHFNC---NKNAS---GASLYLYIPAGELKNLPFGIWDATYKLR 178
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
121 -----ILGVSVDACGVINNNMAAACQWGRSLKLRIPSEELAKIPTSGTWKATVLDY 173
  |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```



```

2190 NAME100001100ANCYVDMCFIDUOISINSSBUEINFUNNFNSD GNFILNAINNDIN 2/3
      : | : | : :: : ||: : || | : ||

```

DD 1060 GNCVDEILILIAVNDP-SNMFVAGIYNFI 1106

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 19, 2003, 09:36:34 ; Search time 44 Seconds
(without alignments)
1302.279 Million cell updates/sec

Title: US-09-839-894-10

Perfect score: 1886
Sequence: 1 MKRKILFIETFFSSVLTFA.....EAGQYMGINVTFTSSQTL 361

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

1: A.Geneseq_19Jun03:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match	Length	ID	Description
1	1886	100.0	361	23	AAM50343
2	927.5	49.2	364	20	AAV22326
3	275.5	14.6	359	22	AAV45919
4	119.5	6.3	770	21	AAV50814
5	117.5	6.2	1386	24	ABR82570
6	117.5	6.2	1449	24	ABR82570
7	114	6.0	650	15	AAV47575
8	111.5	5.9	1005	21	AAV01833
9	111.5	5.9	1011	21	AAV01832

10	110.5	5.9	1529	14	AAV41732
11	110.5	5.9	1601	18	AAV30292
12	109	5.8	2383	21	AAV15945
13	107	5.7	1095	21	AAV01835
14	107	5.7	1101	21	AAV01834
15	106	5.6	671	17	AAV85290
16	106	5.6	1222	21	AAV01830
17	106	5.6	1228	21	AAV01828
18	105	5.6	598	23	ABV47313
19	104.5	5.5	1440	23	ABV54801
20	101	5.4	1358	14	AAV41731
21	101	5.4	1598	18	AAV30291
22	100.5	5.3	1004	21	AAV01841
23	100.5	5.3	1010	21	AAV01840
24	99.5	5.3	969	21	AAV01827
25	99.5	5.3	975	21	AAV01826
26	99	5.2	839	23	ABV56002
27	99	5.2	839	23	ABV98211
28	99	5.2	839	24	ABV66267
29	99	5.2	992	21	AAV01843
30	99	5.2	998	21	AAV01842
31	97	5.1	418	21	AAV30907
32	97	5.1	454	21	AAV30906
33	97	5.1	957	21	AAV01839
34	97	5.1	963	21	AAV01838
35	96	5.1	430	21	AAV37667
36	96	5.1	513	21	AAV39277
37	96	5.1	631	21	AAV39276
38	96	5.1	683	21	AAV39275
39	95.5	5.1	825	21	AAV25547
40	95	5.0	867	23	AAM50341
41	95	5.0	2334	23	ABV31849
42	94.5	5.0	567	20	AAV37170
43	94.5	5.0	665	22	ABV68244
44	94.5	5.0	2353	17	AAV93933
45	94.5	5.0	2411	21	AAV23860

ALIGNMENTS

RESULT 1	
AAM50343	
ID	AAM50343 standard; Protein; 361 AA.
AC	AAM50343;
XX	
DT	18-FEB-2002 (first entry)
XX	
DE	ETEC CS4 pilus CsaE tlp associated protein.
XX	
KW	CS4 pilus; enterotoxigenic; ETEC; csa operon; CsaA; fimbrial;
XX	vacuole; diarrhoea; antibacterial; antidiarrhetic.
OS	Escherichia coli.
XX	
FH	Key
FT	Peptide
FT	Location/Qualifiers
FT	1..23
FT	/label= Signal_peptide
FT	24..361
FT	/label= Mature_protein
XX	
PN	W0200181582-A2.
XX	
PD	01-NOV-2001.
XX	
PF	20-APR-2001; 2001WO-US12914.
XX	
PR	20-APR-2000; 2000US-198666P.
XX	
PA	(UYMA-) UNIV MARYLAND BALTIMORE.
XX	
PI	Altbaum Z, Levine MM, Barry EM,

High molecular weight
Non-tyeable Haemo
E. coli proliferat
Haemophilus influe
Haemophilus influe
Streptococcus faec
H. influenzae stra
Haemophilus influe
Listeria monocytog
Lactococcus lactis
High molecular weight
Non-tyeable Haemo
Haemophilus influe
Haemophilus influe
Chlamydia psittaci
C. psittaci protei
Haemophilus influe
Haemophilus influe
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Haemophilus influe
Eucaalyptus grandis
Human kinase, MEK1
Amnio acid sequenc
Drosophila melanog
Haemophilus adhesi
Haemophilus influe

XX WPI: 2002-049280/06.
DR N-PSDB; AAI10763, AAI10780.
XX
PT New nucleotide sequence, useful as immunogenic agent for generating
PT Immune response against recombinant product of the operon, comprising
PT csa operon which encodes enterotoxigenic *Escherichia coli*-CS4 pili
XX
PS
XX
Claim 10; Page 59; 81bp; English.

CC The present sequence is that of the tip associated protein CsaE
CC of enterotoxigenic *Escherichia coli* (ETEC) strain E11881A. CsaE is
CC encoded by the csaE gene (see AAI10763) of the *E. coli* E11881A csa
CC operon. This operon includes 5 contiguous genes, csaA-csaE, which
CC encode the synthesis of ETEC-CS4 pili. It has been expressed in
CC attenuated *Shigella* strain CVD1204 gnaBA, constructing the *Shigella*
CC expressing Csa4 fimbriae vaccine strain CVD1204 (pGA2-CS4). The
CC CsaE protein has a calculated mol. wt. of 40102.4 and a theoretical
CC pI of 8.74. It shows homology to similar proteins from other ETEC
CC fimbriae. Recombinant CsaA-CsaE polypeptides are used in claimed
CC immunogenic compositions to generate an immune response in a
CC subject. These prevent ETEC colonisation, and hence protect
CC against diarrhoea.

[illegible]

Dd		301 DNVDKAAIR-----PVLPLGQRQARVCPVPLLTLLTOPPNIEKKSGEVGILTYTMMLG	355
OY	358 SQT	360	
	:	:	
Dd	356 TOT	358	
RESULT 4			
ID	AAV50814	standard; Protein; 770 AA.	
AC	AAV50814;		
XX	17-FEB-2000	(first entry)	
DT			
DE	D. melanogaster acetyl-choline receptor protein from clone Da7.		
KW	Acetyl-choline receptor; nicotinic; insect; insecticide; screening;		
KM	neurotransmission; plant protection agent; conductance; AChR.		
XX	Drosophila melanogaster.		
OS			
PN	DEL9819829-AI.		
XX	11-NOV-1999.		
PD			
XX	04-MAY-1998;	98DE-1019829.	
PF			
PR	04-MAY-1998;	98DE-1019829.	
XX	(FARB) BAYER AG.		
PA			
PI	Adamczewski M, Oellers N, Schulte T;		
XX	WP1: 2000-014207/02.		
DR	N-PSSD; AAZ24475.		
PT	New nucleic acid encoding a nicotinic acetylcholine receptor from		
PS	insects, used to identify potential insecticides		
XX	Example 1a; Page 12-14; 26pp; German.		
XX	This invention describes a novel nucleic acid (NA) encoding a nicotinic		
CC	acetyl-choline receptor (I) from insects which can be used as an		
CC	insecticide. Inhibitors of (I) interfere with neurotransmission. (I)		
CC	(also vectors containing it, its regulatory regions, and antibodies		
CC	directed against (I)-encoded proteins) are used to screen for: (a) plant		
CC	protection agents that alter conductance of AChR, potentially useful as		
CC	insecticides, or (b) genes which encode polypeptides that are involved in		
CC	formation of functionally related AChR in insects. (II) are also used to		
CC	isolate and characterize the specified regulatory regions and for		
CC	recombinant production of (II). This sequence represents an		
CC	acetyl-choline receptor isolated from Drosophila melanogaster.		
XX			
SQ	Sequence	770 AA:	
Query Match	6.3%;	Score 119.5;	DB 21; Length 770;
Best Local Similarity	20.4%;	Pred. NO. 0.048;	
Matches	83; Conservative	48; Mismatches	140; Indels 135; Gaps 20;
OY	9 TLFFSVSLTPRAVSADKIPEGDESTINIFGPRDRNESSRKHNILNNH-----	54	
	: : : : :		
Dd	249 TIATISYSGSFMAOLKNSSSSSSSN-----SSNSTSIOILNGINKHSWTFLLIYLNLSSAK	304	
OY	55 --ITAVSESHTLRYBMFLCLSSHNTLGGACPTSENPPSSSVSGSTNTTTQFTERSRLIK	112	
	: : : : :		
Dd	305 VCLAGHGHEKRLIHD-----LDLPYTTERPVYNESDPQLDSGLTLMQIIDVERNQQLIV	359	
OY	113 RELQIGKYQKOLLFRKSVNCPGLTNSAFNCNKNAASGASYLYIPAGEELKNLPGGIWD	172	
	: : : : :		
Dd	360 TNVWLK-----LENNDMKLNRNTSDYGVK-----DLR-IIPHRIWK	395	
OY	173 ATLKLRAVKRRYSER---YGTYYTINTITIKTDKGNIQIWLPO--FKSDARVDLNLRPGG	226	

ID	AAE01833	standard; Protein; 1005 AA.
XX	AAE01833;	
XX	11-SEP-2000	(first entry)
DT		
XX	Haemophilus influenzae strain K21	mature HMW2A protein, SEQ ID NO:41.
XX		
KW	Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;	
KW	non-tyeable Haemophilus influenzae; NTHi; non-encapsulated;	
KW	recombinant production; Escherichia coli; antibacterial; vaccine;	
KW	human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;	
KW	detection; diagnosis.	
XX		
OS	Haemophilus influenzae strain K21.	
XX		
PN	WO200020609-A2.	
XX		
XX	13-APR-2000.	
PF	07-OCT-1999;	99MO-CA00938.
XX		
PR	07-OCT-1998;	98US-0167568.
XX		
PR	08-DEC-1998;	98US-0206942.
XX		
PA	(CONN-) CONNAUGHT LAB LTD.	
PI		
PI	Loosmore SM, Yang Y, Klein MH;	
DR	WPI: 2000-303789/26.	
DR	N-PSDB; AAA52182.	
PT		
PT	Nucleic acid molecule for producing recombinant high molecular weight	
PT	proteins of Haemophilus which are used as a vaccine to provide	
PT	protection against Haemophilus induced diseases in humans -	
XX		
XX	Claim 8; Fig 21A-C; 307pp; English.	
XX		
CC	The invention relates to the recombinant production of Haemophilus	
CC	Influenzae high molecular weight (HMW) proteins in Escherichia coli. The	
CC	expression construct used to effect recombinant expression comprises a	
CC	promoter functional in E. coli (e.g., the 77 promoter) operably linked	
CC	to a modified hmwABC operon from a non-tyeable (non-encapsulated) H.	
CC	Influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene	
CC	clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,	
CC	hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins	
CC	and the hmwB and hmwC genes encode accessory proteins which are	
CC	responsible for post-translational processing and secretion of the HMWA	
CC	proteins. The modified hmwABC operon used in the expression construct of	
CC	the invention contains an A gene modified such that it encodes only the	
CC	mature HMWA. The invention also discloses hmwA genes (AAA52175-52198)	
CC	and HMWA proteins (AAE01824-B01849) from the non-tyeable H. Influenzae	
CC	strains 20yc, KI, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and	
CC	vectors are used for the production of recombinant H. Influenzae HMW	
CC	proteins which can be used as vaccines to mediate a humoral or	
CC	cell-mediated immune response to provide protection against diseases in	
CC	humans caused by H. Influenzae (e.g., otitis media, epiglottitis,	
CC	pneumonia and tracheobronchitis). The HMW proteins are also useful as	
CC	antigens in immunoassays for detecting antidiodes against Haemophilus,	
CC	HMW proteins and/or HMW peptides. The nucleotide sequences encoding the	
CC	HMW proteins can be used to isolate and clone hmw genes from other	
CC	non-tyeable strains of Haemophilus via hybridisation reactions. The	
CC	present sequence represents a mature HMWA protein from a non-tyeable	
CC	strain of H. Influenzae.	
XX		
SO	Sequence	1005 AA;
XX		
Query Match	5.9%;	Score 111.5; DB 21; Length 1005;
Best Local Similarity	20.3%;	Pred. No. 0.41;
Matches	68; Conservative	65; Mismatches 147; Indels 55; Gaps 14

Db	36	EKNAFISTHNLITLGGVNTLGGENSSNFKGINININSKANVTLQAHAGTSHLDKE----	418
Oy	119	GYQLLEFKSYNCPSGITL--NSAHFNCKNAASGSLYLTPPAEL-KNLPFGSIW--DA	173
Db	419	--FTLLTLGNVSVGGINNIIGSNAHLDGNLSIAESAKF-----OGKTNNNINITGTFYNN	471
Oy	174	TLKLRPKRRSEYEGYGVYITITIKLTD-----KGI-----OIWLPOFSARVD	218
Db	472	TADINIKQGVYKQSGITNNGNLNTITNASVQKTIINGNTNNKGLDNIKDIKANAEIQ	531
Oy	219	L--NLRPGGGYIYIGNSVDMCFYDGYSTNSSLEIRFDNNPKSD--GKFLRKINDYK	275
Db	532	IGNISQKEGNLITLISDKINI-----TKRIEIKADTQGDGSDGVASMANLTIKTK	582
Oy	276	EIAVYTLISLLAG---KSLPTNGTSLINTADASLETNNRTTAYTAPBISPVLCWPGRL	332
Db	583	ELVLTDLNLINSGKNAKEITAKNDSLLIGKASSDNSNAKQITFEKVKADSKIS--AGNHV	640
Oy	333	QLDAKYENPEA-----GOYMGINIVTFPPSSQTL	361
Db	641	TLMSKYETSNSDGSSTGSGSDDNIGLITISAKOVY	675
RESULT 9			
AAB01832			
ID	AAB01832	standard; Protein; 1011 AA.	
XX	AAB01832;		
XX	11-SEP-2000	(first entry)	
DE	Haemophilus influenzae strain K21	HMW2A protein, SEQ ID NO:39.	
XX	HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;		
KW	non-typable Haemophilus influenzae; NTNH; non-encapsulated;		
KM	recombinant production; Escherichia coli; antibacterial; vaccine;		
KW	human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;		
KM	detection; diagnosis.		
XX	Haemophilus influenzae strain K21.		
OS	WC200020609-A2.		
PN	13-APR-2000.		
PD	07-OCT-1999;	99MO-CA00938.	
PF	07-OCT-1999;	99MO-CA00938.	
XX	07-OCT-1998;	98US-0167568.	
PR	08-DEC-1998;	98US-0206942.	
XX	(CONN-) CONNAUGHT LAB LTD.		
PA	Loosmore SM, Yang Y, Klein MH;		
XX	WPI: 2000-303789/26.		
DR	N-PSDB: AAA52181.		
XX	Nucleic acid molecule for producing recombinant high molecular weight		
PT	proteins of Haemophilus which are used as a vaccine to provide		
PT	protection against Haemophilus induced diseases in humans -		
XX	Claim 12; Fig 21A-O; 307pp; English.		
CC	The invention relates to the recombinant production of Haemophilus		
CC	influenzae high molecular weight (HMW) proteins in Escherichia coli. The		
CC	expression construct used to effect recombinant expression comprises a		
CC	promoter functional in E. coli (e.g., the λ promoter) operably linked		
CC	to a modified hmwABC operon from a non-typable (non-encapsulated) H.		
CC	influenzae (NTNH). Most HMW-expressing NTNH strains contain two hmw gene		
CC	clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,		
CC	hmwB and hmwC genes. The hmwA genes encode the structural HMW proteins		
CC	and the hmwB and hmwC genes encode accessory proteins which are		
CC	responsible for post-translational processing and secretion of the HMW		

CC proteins. The modified hmwaBC operon used in the expression construct of
 CC the invention contains an A gene modified such that it encodes only the
 CC mature HMWA. The invention also discloses hmwa genes (AA5175-A52198)
 CC and HMWA proteins (AA01824-B01849) from the non-typable H. influenzae
 CC strains J01, K1, K21, LDC2, PMH1, 15 and 12. The nucleic acids and
 CC vectors are used for the production of recombinant H. influenzae HMW
 CC proteins which can be used as vaccines to mediate a humoral or
 CC cell-mediated immune response to provide protection against diseases in
 CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
 CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
 CC antigens in immunoassays for detecting antibodies against Haemophilus,
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
 CC HMW proteins can be used to isolate and clone hmw genes from other
 CC non-typable strains of Haemophilus via hybridisation reactions. The
 CC present sequence represents an HMWA protein from a non-typable strain of
 CC H. influenzae.

XX Sequence 1011 AA;

Query Match 5.9%; Score 111.5; DB 21; Length 1011;

Best Local Similarity 20.3%; Pred. No. 0.41; Mismatches 147; Indels 55; Gaps 14;

Matches 68; Conservative 65; Mismatches 147; Indels 55; Gaps 14;

60 ESHTLYDRMTFLCLSHNTLNGACPTSENPSSSVSGETNITLQTEKRS-LIKRELQIK 118

369 EKNAIFSTHNLTLGNTVLGGENSSNIKGININSKANVTLOAHAGTSHLDKE---- 424

119 GYKQLLEKSVNCSGTL--NSAHFNCNKNMAAGSLYLITPAGEL-KNLPFGIW--DA 173

425 -RTLLGVNVSVCNNILIGSNHNDGNSIAESAF-----QKTNNNINITGTETNNG 477

174 TTKLRVRRYSERYGYTITITIKTD-----KGNL-----QIWLPOKSDARVD 218

478 TADINIKQGVVKKQGITNNGNLITNMSVNGKTIINGNITKKODLNKDKANAELQ 537

219 L--NRPFGGTYIGNSVDMCFYDGYSTNSSLERFODNPKSD-GKFLYLRKINDDT 275

538 IGCNIOKSGNLTISDKNL-----TKRIETKADPDQNSDGSVANSNANLTIKTK 588

276 ELAVTSLLLAG---KSLPTNGTSLNIDAALETNMTNRTAVTPEISVPLWCGR 332

589 ELTLTDLNLSGCKNKAETIAKNSDLILIGKASDONSNAKQITFDKVKDSKIS--AGNHNY 646

333 OLDKAVENPEA-----GOYMGNIIVTFPPSSQTL 361

647 TLNSKVETNSDGSSTGSDNNIGLTLISAKDVTY 681

RESULT 10

AA41732

AA41732 standard; Protein; 1529 AA.

AA41732:

25-MAR-2003 (updated)

26-APR-1994 (first entry)

High molecular weight protein 4 (HMW4).

HMW; high molecular weight protein; virus; vaccine; influenza;

epitope; immunity; haemophilus influenzae.

Haemophilus influenzae.

MO9319090-A1.

30-SEP-1993.

16-MAR-1993; 93WO-US02166.

16-MAR-1992; 92GB-0005704.

(BARE/) BARENKAMP S J.

PA (INRM) INSERM INSR NAT SANTE & RECH MEDICALE.

XX Barenkamp SJ;

XX WPI; 1993-320683/40.

DR N-PDB; AA049511.

PT High molecular weight surface proteins - of non-typable

haemophilus which exhibit immunogenic properties

PS Claim 6; Figure 10; 100pp; English.

CC The isolation and purification of the high molecular weight protein

enables the identification of the major protective epitopes of the

protein by conventional epitope mapping. These epitopes can then be

synthesised using standard techniques and incorporated into fully

synthetic or recombinant vaccines.

(Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 1529 AA;

Query Match 5.9%; Score 110.5; DB 14; Length 1529;

Best Local Similarity 20.7%; Pred. No. 0.95; Mismatches 127; Indels 87; Gaps 18;

Matches 72; Conservative 62; Mismatches 127; Indels 87; Gaps 18;

48 HNTLNHITAYSSHTLYDRMTFLCLSHNTLNGACPTSENPSS-----SVGETNITL 102

744 YNEYSKH--ALNSHNL-----TILGANTLGG-----ENSSSITGINITNKANVTL 790

103 QF-----TEKRSLIKRELQIKGYKQLLEKSVNCPGSLTLNSAHFNCNKNMAAGSLY 154

791 QADTSNMTGLKRRITLGNISVEGNLSLGANANINGNLSI--AEDSTFKGAS----- 843

155 LYTPAGEIKLPPGIIW--DATLKLVRKRYSETGYT-----INTTIKTL----- 199

844 -----DLNITGFTNNGTANINIKQGVVLOGDIINKGILNITNNASTQKTTING 895

200 ----DKGNIOIWLPOKSDARVDL--NLRPTGGGTYIGNSVDMCFYDGYSTNSSLER 253

896 NITNEKGDNL--KNIKADAEIOLGNSIGKEBNLTISDKVNI-----TQGITKAG 946

254 FQ-----DNPKSGKTYLRKINDTKEIAYTSLLAG---KSLPTNGTSLNIDAA 306

947 VEGGRSDSSAEANANLTIQ-----TEKELAGDLNLSGFKAKETIAKNSGDLTIGNASG 1001

307 ETMWNRTIANTPEISVPLWCWG--RLQOLDKAVENPEAGGYMGNIIVT 353

1002 NADAKR---VTFDKVKDSKISTDGHNVTLNSEVKTSGSSNAGNDNST 1046

RESULT 11

AAW30292

AAW30292 standard; Protein; 1601 AA.

AAW30292:

14-APR-1998 (first entry)

Non-typable Haemophilus high mol.wt. surface protein HMW4.

Non-typable Haemophilus; high molecular weight surface protein;

HMW4; immunogen; vaccine; otitis media.

Haemophilus influenzae strain 5.

Key Location/Qualifiers

Misc-difference 372 /note- "encoded by TCR"

Misc-difference 400 /note- "encoded by AAT"

MO9736914-A1.

Db 1080 GNGVDETTLATYKDP-SHHPVAGITVNFT 1108

RESULT 13

AA01835
ID AAB01835 standard; Protein; 1095 AA.
AC AAB01835;
XX
DT 11-SEP-2000 (first entry)

DE Haemophilus influenzae strain LCD2 mature HMMA protein, SEQ ID NO:45.
XX
XX Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KM non-typhable Haemophilus influenzae; NTH1; non-encapsulated;
KM recombinant production; Escherichia coli; antibacterial; vaccine;
KM human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
XX detection; diagnosis.

OS Haemophilus influenzae strain LCD2.
MO200020609-A2.
PD 13-APR-2000.

PF 07-OCT-1999; 99WO-CA00938.
XX
PR 07-OCT-1998; 98US-0167568.
XX 08-DEC-1998; 98US-0206942.
XX
PA (CONN-) CONNAUGHT LAB LTD.

PI Loomore SM, Yang Y, Klein MH;
XX
DR WPI; 2000-303789/26.
XX N-PSDB; AAAS2184.

PT Nucleic acid molecule for producing recombinant high molecular weight
XX proteins of Haemophilus which are used as a vaccine to provide
XX protection against Haemophilus induced diseases in humans -

PS Claim 8; Fig 22A-P; 307pp; English.

XX
XX The invention relates to the recombinant production of Haemophilus
CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC expression construct used to effect recombinant expression comprises a
CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC to a modified hmwABC operon from a non-typhable (non-encapsulated) H.
CC influenzae (NTH1). Most HMW-expressing NTH1 strains contain two hmw gene
XX clusters termed hmwIABC and hmw2ABC. Each hmwABC operon comprises hmwA,
XX hmwB and hmwC genes. The hmwA genes encode the structural HMMA proteins
XX and the hmwB and hmwC genes encode accessory proteins which are
CC responsible for post-translational processing and secretion of the HMMA
CC proteins. The modified hmwABC operon used in the expression construct of
CC the invention contains an A gene modified such that it encodes only the
CC mature HMMA. The invention also discloses hmwA genes (AAAS2175-AS2198)
CC and HMMA proteins (AAB01824-B01849) from the non-typhable H. influenzae
CC strains Joyce, K1, K21, LCD2, PMH1, 15 and 12. The nucleic acids and
CC vectors are used for the production of recombinant H. influenzae HMW
CC proteins which can be used as vaccines to mediate a humoral or
CC cell-mediated immune response to provide protection against diseases in
CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC antigens in immunoassays for detecting antibodies against Haemophilus,
CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC HMW proteins can be used to isolate and clone hmw genes from other
CC non-typhable strains of Haemophilus via hybridisation reactions. The
CC present sequence represents a mature HMMA protein from a non-typhable
CC strain of H. influenzae.

XX Sequence 1095 AA;

Query Match 5.7%; Score 107; DB 21; Length 1095;
Best Local Similarity 24.6%; Pred. No. 1.2;
Matches 71; Conservative 44; Mismatches 110; Indels 64; Gaps 15;

QY 45 SPKHNILNHTITAYSSSHLYD-----RMFLCLSSHTINTLNGACPSENSSSS 93
DB 281 SIRKPIVSN--VHDSNHTLFNGNSVVLGGDVNFHFNASSSNHWGVIYSQNMASE 337
QY 94 VSGETNITLOFTEKRLKRELOIKGYKOLKFSVNCPSGLTLNSAHFNCKNKAASGAL 153
DB 338 GS-----SLRF-----KSGESTTAF---TIESDLTNATGNSISLNOVAGIDG 378
QY 154 VLYTPAGELKLPFGGIWDATLKLRYKRYSEYGYTYT---NITIKLTDKNGIQIWP 209
DB 379 NLOKSLVANKNITFEES--GNITLADKKPIEIKGNTYKEGANVTLRGANYGNDKSAI- 434
QY 210 QEKSDARVLNLRPTGGGTYIGRN-----SDMKCFDGIYSTNSSLEIFPDNNRPSD-- 262
DB 435 SIRGANTNKGNTLVGSAINIEKNLVEGSAFLANPNYSFVNSGL---FDNOKGSNIS 490
QY 263 ---GKFLRKINDTKEIAYTIS-----LLAGKSLPTNGTSLNIAD 302
DB 491 TAKGAHFEDIN-NTKSLNITNDSAYRTIIEG-NITNSNG-DLNIITD 536

RESULT 14

AA01834
ID AAB01834 standard; Protein; 1101 AA.

AC AAB01834;

DT 11-SEP-2000 (first entry)

DE Haemophilus influenzae strain LCD2 HMMA protein, SEQ ID NO:43.

XX
XX HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KM non-typhable Haemophilus influenzae; NTH1; non-encapsulated;
KM recombinant production; Escherichia coli; antibacterial; vaccine;
KM human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
XX detection; diagnosis.

OS Haemophilus influenzae strain LCD2.

PN MO200020609-A2.

PD 13-APR-2000.

PF 07-OCT-1999; 99WO-CA00938.

XX
XX 07-OCT-1998; 98US-0167568.
XX 08-DEC-1998; 98US-0206942.

PA (CONN-) CONNAUGHT LAB LTD.

PI Loomore SM, Yang Y, Klein MH;

XX
XX WPI; 2000-303789/26.
XX N-PSDB; AAAS2183.

PT Nucleic acid molecule for producing recombinant high molecular weight
XX proteins of Haemophilus which are used as a vaccine to provide
XX protection against Haemophilus induced diseases in humans -

PS Claim 12; Fig 22A-P; 307pp; English.

XX
XX The invention relates to the recombinant production of Haemophilus
CC influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC expression construct used to effect recombinant expression comprises a
CC promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC to a modified hmwABC operon from a non-typhable (non-encapsulated) H.
CC influenzae (NTH1). Most HMW-expressing NTH1 strains contain two hmw gene
XX clusters termed hmwIABC and hmw2ABC. Each hmwABC operon comprises hmwA,
XX hmwB and hmwC genes. The hmwA genes encode the structural HMMA proteins

CC and the hmwB and hmwC genes encode accessory proteins which are
 CC responsible for post-translational processing and secretion of the HMW
 CC proteins. The modified hmwABC operon used in the expression construct of
 CC the invention contains an A gene modified such that it encodes only the
 CC mature HMW. The invention also discloses hmwA genes (AA52175-45218)
 CC and HMW proteins (AA801824-B01849) from the non-typable H. influenzae
 CC strains Joyce, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
 CC vectors are used for the production of recombinant H. influenzae HMW
 CC cell-mediated immune response to provide protection against diseases in
 CC humans caused by H. influenzae (e.g., otitis media, epiglottitis,
 CC pneumonia and tracheobronchitis). The HMW proteins are also useful as
 CC antigens in immunoassays for detecting antibodies against Haemophilus,
 CC HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
 CC HMW proteins can be used to isolate and clone hmw genes from other
 CC non-typable strains of Haemophilus via hybridisation reactions. The
 CC present sequence represents an HMW protein from a non-typable strain of
 H. influenzae.

Sequence 1101 AA:

Query Match 5.7%; Score 107; DB 21; Length 1101;

Best Local Similarity 24.6%; Pred. No. 1.2; Mismatches 110; Indels 64; Gaps 15;

Matches 71; Conservative 44; Mismatches 110; Indels 64; Gaps 15;

QY 45 SPKHILNNHITAYSESHLYD-----RMIFCLSSHNTLNGACPTSENPSSSS 93
 DB 287 SIKRPIVSN---VHDGNHTLFENGNAVSLGGDVNFHFNASSNMHTHGVIKSNQFNASE 343
 QY 94 VSGENITLOFTEKRSILKRELQIKGYQLFKSYVNCPSGLTNSAHFNCNKAASGASL 153
 DB 344 GS-----SLRF-----KSEGSTYTA---TIESDLTLNATGNSISLNOVAGIDG 384
 QY 154 YLYIPAGELKNLPFGGIMDATLKLVRKRYSEYGTYI---NITIKLTDGNIQIOWP 209
 DB 385 NLQKSLVANKNKNTFEG---GNITLADKKPIEIKGNITVKEGANTLRKANYGNDKSKAL- 440
 QY 210 QPKSARDLNRPFPGGGTYIGRN-----SYDMCFYDGYSTNSSSLEIRFODNPKSD-- 262
 DB 441 SIRGVNTKGNLTIVGSAINIEKRLTVEGSAKFLANPNISFVNSGI---FPOCKSKNIS 496
 QY 263 ---GKFLYLRKINDPTEIAYTLSS-----LLAGKSLPPTNGTSINIA 302
 DB 497 IAKGGAHFRDIN-NTRKSLNITTNSDSAYTITEG-NITNSNG-DLNIITD 542

ULN 15
 85290

AA85290 standard; Protein; 671 AA.

AA85290;

02-APR-1996 (first entry)

Streptococcus faecalis autolysin.

Lysin; autolysis; culture; lactic acid bacteria; fermentation;
 cheese; foodstuffs; induction.

Streptococcus faecalis.

W09531561-A1.

23-NOV-1995.

12-MAY-1995; 95WO-NL00170.

12-MAY-1994; 94EP-0201353.

(UNIL) QUEST INT BV.

Buist G, Kok J, Ledebuer AM, Venema G;

DR WPI; 1996-010946/01.
 XX Lysis of a culture of lactic acid bacteria in, e.g. cheese
 PT production - by in situ prodn. of an auto-lysin, regulated by an
 PT inducible promoter.

PS Disclosure; Page 66-69; 103pp; English.

CC In situ production of a homologous autolysin or a heterologous
 CC autolysin from a food grade gram positive bacteria, can be used in
 CC a process for the lysis of a culture of lactic acid bacteria. The
 CC process can be used in the manufacture of products containing
 CC cultures of lactic acid bacteria e.g. cheese, where the culture is
 CC lysed following the completion of fermentation. The enhanced
 CC induction of the autolysin is performed some hours after the
 CC fermentation is finished. No extra lysis needs to be added and the
 CC lysis does not need to be isolated or encapsulated. The time of
 CC lysis can be precisely controlled. This is the Streptococcus
 CC faecalis autolysin.

SQ Sequence 671 AA:

Query Match 5.6%; Score 106; DB 17; Length 671;

Best Local Similarity 23.6%; Pred. No. 0.74; Mismatches 103; Indels 94; Gaps 17;

Matches 73; Conservative 39; Mismatches 103; Indels 94; Gaps 17;

QY 41 RNESSPKHNT-LNNHITAYSESHLYDMMTFCLSSHNTLNGACPTSENPSSSV----- 94
 DB 309 RYATDPYNAKLNNTITAY--NLTOYDPPSSGNGTGGGTGVNPGTGSNNQSTNTTYIVK 366
 QY 95 SGEI---NITLOP-----TEKRSI--IKRELQIKGYQLFKSYVNCPSGLTNSAHFNCNKN 146
 DB 367 SCDTLNKIAAOYGVSVANLRSMWNGISGDLIFVGOKLIYKKA---SGNTGGSGNGSGNNN 423
 QY 147 AASGASLYIIPAGELKNLPFGGIMDATLKLVRKRYSEYGTYIINITIKLTDGNIQI 206
 DB 424 -QSGINTYTYVKSGLTN-----KIAQIC-----VTV-----ANLRS 455
 QY 207 WLPOKSDARVDLNI-----RPTGGGTYYIGRNSVDMCFYDGYSTNSSSLEIRFODNN 258
 DB 456 W-----NGISGDLIFVGOKLIYKKGTSNT-----GGSSNGS-----NNN 491
 QY 259 PKSDGKTYLRKINDPTEIAYTLSS-----LLAGKSLPPTNGTSINIA 303
 DB 492 QSGINTYTYIKSGDPLNKIAAOYGVSVANLRSMWNGISGDLIFAGOKIIVKKGTSNGTGS 551
 QY 304 ASLETNMNR 312
 DB 552 SMGGSNNQ 560

Search completed: August 19, 2003, 09:45:52
 Job time : 46 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 16:04:22 ; Search time 3990 Seconds

(without alignments)
11134.801 Million cell updates/sec

Title: US-09-839-894-9

Perfect score: 1086
Sequence: 1 atgataagaattattat.....caagtagtcaaacactctag 1086

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Immun DB seq length: 0
Immun DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_da: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_da: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_htg_hum: *
31: em_htg_in: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pln: *
35: em_htg_rod: *
36: em_htg_mam: *
37: em_htg_vrt: *
38: em_sy: *
39: em_htgo_hum: *
40: em_htgo_mus: *
41: em_htgo_other: *

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1086	100.0	1086	AX352333	AX352333 Sequence
2	1086	100.0	7239	AF296132	AF296132 Escherich
3	1086	100.0	7239	AX352351	AX352351 Sequence
4	1086	100.0	7239	AX741418	AX741418 Sequence
5	852.4	78.5	7174	1 ECCOFPA1A	AX741418 Sequence
6	310.8	28.6	5336	6 ECCOCD	M55661 E.coli CFA/
7	310.8	28.6	5336	6 AX741413	AX741413 Sequence
8	309.6	27.5	6956	1 AX716491	AX716491 Escherich
9	298.8	27.5	5798	1 ECCOTABCD	AX741415 Sequence
10	298.8	27.5	5798	6 AX741415	AX741415 Sequence
11	63.4	5.8	1141	6 AX083744	AX083744 Sequence
12	57.6	5.3	61864	3 CEY50E8A	AL117200 Caenorhab
13	53.6	4.9	5230	8 AF253047	AF253047 Candida a
14	50.6	4.7	29143	3 AC115594	AC115594 Dictyoste
15	50.6	4.7	88549	3 AC116979	AC116979 Dictyoste
16	50.2	4.6	153477	2 AC006278	AC006278 Dictyoste
17	50.2	4.6	251551	3 AE014844	AE014844 Plasmodiu
18	50.2	4.6	2000	6 AX655393	AX655393 Sequence
19	49.6	4.6	253924	3 AE014822	AE014822 Plasmodiu
20	49.4	4.5	10279	6 AX279996	AX279996 Sequence
21	49.4	4.5	10279	6 AX346493	AX346493 Sequence
22	49.4	4.5	10279	6 AX356452	AX356452 Sequence
23	49.2	4.5	253151	3 AE014842	AE014842 Plasmodiu
24	48.6	4.5	1141	6 AX083744	AX083744 Sequence
25	48.2	4.4	9504	6 AX323793	AX323793 Sequence
26	48.2	4.4	349980	6 AX344570	AX344570 Sequence
27	48.2	4.4	3549	3 AF310884	AF310884 Dictyoste
28	47.6	4.4	169794	2 AC004588	AC004588 Plasmodiu
29	47.6	4.4	196149	2 AC004709	AC004709 Plasmodiu
30	47.6	4.4	252650	3 AE014847	AE014847 Plasmodiu
31	47.2	4.3	12940	3 AE001386	AE001386 Plasmodiu
32	46.8	4.3	105682	3 AC116957_3	Continuation (4 of
33	46.6	4.3	110000	2 PFMA13_10	Continuation (11 o
34	46.6	4.3	110000	2 PFMA13_05	Continuation (6 of
35	46.6	4.3	110000	2 PFMA13_09	Continuation (6 of
36	46.6	4.3	132254	3 AC116330	AC116330 Dictyoste
37	46.2	4.3	76568	3 MBREV	AF538053 Monosiga
38	46.2	4.3	110000	3 AC116305_1	Continuation (2 of
39	46.2	4.3	11882	3 AC115612	AC115612 Dictyoste
40	46.2	4.2	153751	3 AC116551	AC116551 Dictyoste
41	46.2	4.2	178376	2 BX323051	BX323051 Danto rer
42	46.2	4.2	236521	2 AC129274	AC129274 Rattus no
43	46.2	4.2	331039	3 AC116988	AC116988 Dictyoste
44	45.6	4.2	251448	3 AE014819	AE014819 Plasmodiu
45	45.2	4.2	51405	2 AC019512	AC019512 Drosophila

ALIGNMENTS

RESULT 1	AX352333	1086 bp	DNA	linear	PAT 06-FEB-2002
LOCUS	AX352333				
DEFINITION	Sequence 9 from Patent WO0181582.				
ACCESSION	AX352333				
VERSION	AX352333.1				
KEYWORDS	GI:18617616				
SOURCE	Escherichia coli				
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Escherichia coli				
REFERENCE	Altpoun,Z., Levine,M.M. and Barry,E.M.				
AUTHORS	Isolation and characterization of the csa operon (etec-csa p11)				
TITLE	and methods of using same				

JOURNAL Patent: WO 0181582-A 9 01-NOV-2001;
University of Maryland, Baltimore (US)
Location/Qualifiers
FEATURES
source

CDS

1. 1086
/organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"
1. 1086
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAD2854.1"
/db_xref="GI:18617617"
/translation="MNKILFLTFEFSVLETFVAVSADKIDGDSINIEGPRDRNES
SPKNIILNNHIAVSESHLYDRMTFLCLSSHNLTNGACPTSENPSSSVGENTIL
QPTKRSLIKRELDIKGKQLLEFKSVNGPSGLTNSACFNKNAASGALYIPAG
ELKNIPEGIWDATILKLVKRYSEYGTIVINTIKLTDKNGIOLIMPOKSAVAD
LNLRTGGGTGYIGNSVDMCPYDGYSTSSSLFTRPDNNKSGKFKRLINDTTE
IAYLSILIAKSLTPNGTSLNADASLETNMNRTIATVMPETISPVLCWPRRLD
DAKVENPEAGYMKGNINVTFTPSQTL"
BASE COUNT 366 a 187 c 202 g 331 t
ORIGIN

Query Match 100.0%; Score 1086; DB 6; Length 1086;
Best Local Similarity 100.0%; Pred. No. 4,3e-223;
Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGATAAGATTTTATTTTATTTTACATTTGTTTCCTCAGTACTTTTACATTTGCT 60
DB 1 ATGATAAGATTTTATTTTATTTTACATTTGTTTCCTCAGTACTTTTACATTTGCT 60
OY 61 GTATCGGAGATTAATTTCCCGGAGATGAAGCATTAATTTTGGCCCGCGTAC 120
DB 61 GTATCGGAGATTAATTTCCCGGAGATGAAGCATTAATTTTGGCCCGCGTAC 120
OY 121 AGGAAGCATCTTCCCAACATATATTTAATATACATATTACAGATACAGTAA 180
DB 121 AGGAAGCATCTTCCCAACATATATTTAATATACATATTACAGATACAGTAA 180
OY 121 AGGAAGCATCTTCCCAACATATATTTAATATACATATTACAGATACAGTAA 180
DB 121 AGGAAGCATCTTCCCAACATATATTTAATATACATATTACAGATACAGTAA 180
OY 181 AGTCACTCTGTATGATAGATGATGATTTTATGTTTGTCTGCACATATACCTTAT 240
DB 181 AGTCACTCTGTATGATAGATGATGATTTTATGTTTGTCTGCACATATACCTTAT 240
OY 181 AGTCACTCTGTATGATAGATGATGATTTTATGTTTGTCTGCACATATACCTTAT 240
DB 181 AGTCACTCTGTATGATAGATGATGATTTTATGTTTGTCTGCACATATACCTTAT 240
OY 241 GGAGATGTCACACAGTGAAGATTCCTAGCTTCATCGGAGCGGGAACAATTA 300
DB 241 GGAGATGTCACACAGTGAAGATTCCTAGCTTCATCGGAGCGGGAACAATTA 300
OY 241 GGAGATGTCACACAGTGAAGATTCCTAGCTTCATCGGAGCGGGAACAATTA 300
DB 241 GGAGATGTCACACAGTGAAGATTCCTAGCTTCATCGGAGCGGGAACAATTA 300
OY 301 ACATTACATTTTACGAAAAAGAGTTTAAATAAAGAGACTACAATTAAGGCTAT 360
DB 301 ACATTACATTTTACGAAAAAGAGTTTAAATAAAGAGACTACAATTAAGGCTAT 360
OY 301 ACATTACATTTTACGAAAAAGAGTTTAAATAAAGAGACTACAATTAAGGCTAT 360
DB 301 ACATTACATTTTACGAAAAAGAGTTTAAATAAAGAGACTACAATTAAGGCTAT 360
OY 361 AAACAATTTATGTTCAAAAAGTTTAACTGCCATCCGCGCTTAACACTTAATCAGCTCAT 420
DB 361 AAACAATTTATGTTCAAAAAGTTTAACTGCCATCCGCGCTTAACACTTAATCAGCTCAT 420
OY 421 TTTAAGTGAATAAAGCGGCGTTCAGGTGCAAGTTTAAATTAATATTCGCGTGGC 480
DB 421 TTTAAGTGAATAAAGCGGCGTTCAGGTGCAAGTTTAAATTAATATTCGCGTGGC 480
OY 421 TTTAAGTGAATAAAGCGGCGTTCAGGTGCAAGTTTAAATTAATATTCGCGTGGC 480
DB 421 TTTAAGTGAATAAAGCGGCGTTCAGGTGCAAGTTTAAATTAATATTCGCGTGGC 480
OY 481 GAACATAAATTTGCTTTGATGATATCGATGCTACTGTAAGTTAAGATAAA 540
DB 481 GAACATAAATTTGCTTTGATGATATCGATGCTACTGTAAGTTAAGATAAA 540
OY 481 GAACATAAATTTGCTTTGATGATATCGATGCTACTGTAAGTTAAGATAAA 540
DB 481 GAACATAAATTTGCTTTGATGATATCGATGCTACTGTAAGTTAAGATAAA 540
OY 541 AGAGATATATGAGACTATGAACTTACACTTAAATATACATTAATTAATGAT 600
DB 541 AGAGATATATGAGACTATGAACTTACACTTAAATATACATTAATTAATGAT 600
OY 541 AGAGATATATGAGACTATGAACTTACACTTAAATATACATTAATTAATGAT 600
DB 541 AGAGATATATGAGACTATGAACTTACACTTAAATATACATTAATTAATGAT 600
OY 601 AAGGGAATATTCAGATATGTTTACCTCAGTTCAAAAGTGCCTCGGTGATCTTAAC 660
DB 601 AAGGGAATATTCAGATATGTTTACCTCAGTTCAAAAGTGCCTCGGTGATCTTAAC 660
OY 601 AAGGGAATATTCAGATATGTTTACCTCAGTTCAAAAGTGCCTCGGTGATCTTAAC 660
DB 601 AAGGGAATATTCAGATATGTTTACCTCAGTTCAAAAGTGCCTCGGTGATCTTAAC 660
OY 661 TTGGCTCAACTGCTGGGCGCATATATTTGAAGAATTTCTGTGATATGCTTTTAT 720
DB 661 TTGGCTCAACTGCTGGGCGCATATATTTGAAGAATTTCTGTGATATGCTTTTAT 720

DB 661 TTGGCTCAACTGCTGGGCGCATATATTTGAAGAATTTCTGTGATATGCTTTTAT 720
OY 721 GATGATATATGATTAACAGCAGCTCTTTGGAGATAGATTTAGATTAACATCTTAA 780
DB 721 GATGATATATGATTAACAGCAGCTCTTTGGAGATAGATTTAGATTAACATCTTAA 780
OY 781 TCTGATGGAAATTTTATCTAAGAAATAAATGATGACACCAAGAAATTTGATATAC 840
DB 781 TCTGATGGAAATTTTATCTAAGAAATAAATGATGACACCAAGAAATTTGATATAC 840
OY 841 TTTGACATCTCTGGGGGTAAGTTTAACTCCACCAAAATGGAAGCTATTAATAT 900
DB 841 TTTGACATCTCTGGGGGTAAGTTTAACTCCACCAAAATGGAAGCTATTAATAT 900
OY 901 GCTGACGAGCTTCTCTGGAACAACAAGTGAATGAATTAACAGCTGCACCATCCAGAA 960
DB 901 GCTGACGAGCTTCTCTGGAACAACAAGTGAATGAATTAACAGCTGCACCATCCAGAA 960
OY 961 ATCAGTGTCCGCTGTGTGTGTGTGGCTGAGAGCTTTCGAATTTGATGCAAAAGTGAAT 1020
DB 961 ATCAGTGTCCGCTGTGTGTGTGTGGCTGAGAGCTTTCGAATTTGATGCAAAAGTGAAT 1020
OY 1021 CCCGAGGCTGACATATATGAGTAAATTAATGTTACTTTTCACACCAAGTACAAACA 1080
DB 1021 CCCGAGGCTGACATATATGAGTAAATTAATGTTACTTTTCACACCAAGTACAAACA 1080
OY 1081 CTCTAG 1086
DB 1081 CTCTAG 1086

RESULT 2
AF296132 7239 bp DNA linear BCT 27-FEB-2003
LOCUS
DEFINITION
Escherichia coli csa operon, complete sequence, and insertion
sequence IS1.
ACCESSION
AF296132
AF296132.1 GI:15419711
KEYWORDS
ORGANISM
Escherichia coli
Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.

REFERENCE
AUTHORS
TITLE
1 (bases 1 to 7239)
Altboum, Z., Levine, M.M., Galen, J.E. and Barry, E.M.
Genetic Characterization and Immunogenicity of Coli Surface Antigen
4 from Enterotoxigenic Escherichia coli when It Is Expressed in a
Shigella Live-Vector Strain
Infect. Immun. 71 (3), 1352-1360 (2003)
JOURNAL
MEDLINE
22483692
PUBMED
12595452

REFERENCE
AUTHORS
TITLE
2 (bases 1 to 7239)
Altboum, Z.D., Levine, M.M., Galen, J.E. and Barry, E.M.
Isolation and characterization of ETEC CS4 fimbriae encoding genes,
and their expression in Shigella flexneri 2a guBAA strain CVD 1204
unpublished
3 (bases 1 to 7239)
Altboum, Z.D., Levine, M.M. and Barry, E.M.
Direct Submission
Submitted (14-AUG-2000) Center for Vaccine Development, University
of Maryland, School of Medicine, 685 W. Baltimore Street,
Baltimore, MD 21201, USA

FEATURES
source
Location/Qualifiers

1. 7239
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="E1881A"
/serotype="O25:H42"
/db_xref="taxon:562"
/note="enterotoxigenic"
283. 6849
/gene="csa operon"
/note="encodes ETEC CS4 fimbrial proteins; CFA/IV"

```

gene      283. .999
CDS       /gene="csaa"
          283. .999
          /gene="csaa"
          /note="colonization factor antigen a; may stabilize the
          major pilin protein"
          /transl_start=1
          /codon_start=1
          /product="periplasmic chaperone-like protein CsaA"
          /protein_id="AK97134.1"
          /db_xref="GI:15419712"
          /translation="MKRLFCGLSLITTPPYANANPMIYPSKDLKNGSELIRVYSKS
          KEIQIKITTKKINPTEHEVDMPNDGCVTPPOVILPAGSKSIRLQFPIP
          KKEELRYVEAVKPDSEKENVINDKLTTELSTNIIALIRSLPSEONISLISRNA
          RKNIIYNGNVAAGVKDIYFCSSNIDSCVAKTHNKNIPKESFDLIVNNFSYVE
          IKLNHEDIEQGLIQKVP"
          1028. .1531
          /gene="csab"
          1028. .1531
          /gene="csab"
          /note="colonization factor antigen b"
          /codon_start=1
          /transl_table=11
          /product="Csa major fimbriae subunit CsaB"
          /protein_id="AK97135.1"
          /db_xref="GI:15419713"
          /translation="MKIKTKIGAMALTFMFVASASAVEKNITVTSVPTDIILOAD
          GSSLPTAVELTSPASREFENKIAKTHVNTNKNVILKVLKATDPTLTNTDSTKOLP
          ITVSYGKRLSTADYTFEPALNFTSGVTGSSSDYIAGTTAAPPNAGTSGVYS
          IIMRTLAS"
          1589. .4192
          /gene="csac"
          1589. .4192
          /gene="csac"
          /note="putative outer membrane protein involved in the
          transport of the major fimbriae subunits across the outer
          membrane; colonization factor antigen c"
          /codon_start=1
          /transl_table=11
          /product="usher protein CsaC"
          /protein_id="AK97136.1"
          /db_xref="GI:15419714"
          /translation="MKKNLTLYTTITAMLTPIYSVSGDIPNSFRDLMGODEYEVKL
          YGQTLGIHRIKTPPIHKKYSPESILDKINVKKEKKLSVETNSFSRGNKNSCGGN
          AIIQYNCNIIKRSVDYIVDDVNVNLEIGNEFLDSEAHNDEYHQLSRVYKAFLOS
          QIINVSDSGKYKLSYSGNSALGIDTSTAYALMMANNYKNFNGSNMERTINLAFRH
          DLKRYIYQGRMDRTDLSOSISGNFNLPLPDIDGIRITGTGSIKINMTDGIASP
          VMYMLTFPSRVEAFRNNOILGYVYIDSGVNEIDTAPLPGYSDIKIKIPENQIAREE
          IIPNKGSSIGDMQDVFIQGNITINDKDRITEKONNHSVNAIGRLPITKINISYO
          OGASVINDKNYIEGSLKWNISGLISLSEFSFLWGDNAKNQOSIYTDGFLSEYH
          NDKRVNCGRNYNAGWSGCEYSASLSIPDLGWTSLGSDYSSEVYKNHILSEYG
          FYNONIYKGRTOQMOLSTSLKMDYINPAIGINSEBOLDKGGYISVLTAS
          RENSINAGSYNYSRGKYSNELFDVGYMTSTNGYHEGRMFPKNRHNABEERLSGR
          INNRFGDLNGSPSMKNKRNMTNSNHSILGYNISFALTSNGPFGWSASAGLTLAGI
          IKKSNDRTKNLKVTGALGXDRSISNDNAFLIPVALPILIEDNNTGKNTISVL
          APLNNDMFLPGNVYVELETKVSVYIGKFGDKGTPISGAILVMEPHYIIDEDEG
          SEFEYGENEKLFLKGRITTYCOLGRNKHKGIVFDVYICDYNSTSSLPDEYKMR
          VQDLAKNDKG"
          4196. .5281
          /gene="csae"
          4196. .5281
          /gene="csae"
          /note="probably found at the fimbriae tip and involved in
          attachment of the bacteri; colonization factor antigen e"
          /codon_start=1
          /transl_table=11
          /product="minor pilin protein CsaE"
          /protein_id="AK97137.1"
          /db_xref="GI:15419715"
          /translation="ANKILFIPTLFSSVLTFFAVSADKIPGDESITINIRPDRNES
          SPKHNTLNHITAYSESHLYDRTFLCLSSHTLNGACPTSPSSSVSGGTNTL
          OPEKRSILKREIQIGYKQLERKSYNCSGLTLNAGHPCNNNAASGSLYLYTAG
          ELKNLPFGCIWDTLKLRYKRRISRTYGTITNTITTKLIDKNGIQIIMLPQFKDARYD

```

```

LNLKPTGGCTTIGRNSVDMCFYDGYSTNSSLSLEIRPDNNPKSGKFLRKINDTYKE
IAVTEILLAGKSLPTPTNGTSLNIADAASLEETWMNRITATVMPBEISVPILCMFGRLQL
DAKVENPAGQVYGNINVTETPSSQTL"
complement(5790. .6119)
/gene="csad"
/note="truncated; CsaD"
/insertion_seq="IS1"
repeat_region 6096. .6870
gene 6143. .6427
/gene="isa1a"
/note="truncated; Insa"
6346. .6849
/gene="isa1b"
6346. .6849
/gene="isa1b"
/codon_start=1
/transl_table=11
/product="Insb"
/protein_id="AK97138.1"
/db_xref="GI:15419716"
/translation="MEGNSPHYGRMFPOHFTSLUKLRPOSYTSRIQPSDYIYCAEMD
EOMGYGAKSRORWLFYADRLAKTVAAHVEGRTAATIGRLMSLSPDYIYMTDGE
WPIYESRLKGLHIVISKRTTORIERNLNLROHLARLGRKLSFSKVELHDVIGHY
LNTRHYO"

BASE COUNT 2436 a 1181 c 1345 g 2277 t
ORIGIN
Query Match 100.0%; Score 1086; DB 1; Length 7239;
Best Local Similarity 100.0%; Pred. No. 3,3e-223;
Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAATAGATTTTATTTATTTTATTTTACATGTTTTCCTTCAGTACTTTTATACATTTGCT 60
DB 4196 ATGAATAGATTTTATTTATTTTATTTTATTTTACATGTTTTCCTTCAGTACTTTTATACATTTGCT 4255
QY 61 GTATCGGACATATAAATTCGCCGAGATGAAGAATATAAATTTTGGCCCGCGTGAC 120
DB 4256 GTATCGGACATATAAATTCGCCGAGATGAAGAATATAAATTTTGGCCCGCGTGAC 4315
QY 121 AGGAGAAATCTTCCCAACAAATATATTTAATTAACATTTTACAGATACATGAA 180
DB 4316 AGGAGAAATCTTCCCAACAAATATATTTAATTAACATTTTACAGATACATGAA 4375
QY 181 AGTCATACCTGTATGATAGATGACATTTTATGTTTGTCTTCACATACATTAAT 240
DB 4376 AGTCATACCTGTATGATAGATGACATTTTATGTTTGTCTTCACATACATTAAT 4435
QY 241 GGAGCATGTCAACCACTGGAATTCCTAGAGTTTCATCGGTCAAGCGGTGAACAAATATA 300
DB 4436 GGAGCATGTCAACCACTGGAATTCCTAGAGTTTCATCGGTCAAGCGGTGAACAAATATA 4495
QY 301 ACATTCATATTTACGAAAAAAGAGTTTAAATAAAGAGCTACAAATTAAGGCTAT 360
DB 4496 ACATTCATATTTACGAAAAAAGAGTTTAAATAAAGAGCTACAAATTAAGGCTAT 4555
QY 361 AAACAATATTGTTCAAAAGCTGTTAACTGCCATCCGCGCTAACACTTAAGTCAAGTCAT 420
DB 4556 AAACAATATTGTTCAAAAGCTGTTAACTGCCATCCGCGCTAACACTTAAGTCAAGTCAT 4615
QY 421 TTTAAGCTATATAAAGCGCGCTTCAAGTGCAGATTTATATTTATATATCTGCTGGC 480
DB 4616 TTTAAGCTATATAAAGCGCGCTTCAAGTGCAGATTTATATTTATATATCTGCTGGC 4675
QY 481 GAACATAAAATTTGCTTTTGTGTGATCTGGAGTCTCTGTAAGTTAAAGTAAAA 540
DB 4676 GAACATAAAATTTGCTTTTGTGTGATCTGGAGTCTCTGTAAGTTAAAGTAAAA 4735
QY 541 AGACGATATAGTGAACCTATGGAACCTATACATATAATATCTATTAATTAAGTAA 600
DB 4736 AGACGATATAGTGAACCTATGGAACCTATACATATAATATCTATTAATTAAGTAA 4795
QY 601 AAGGGAATATTAGATATGTTACTGCACTTAAAGTCAAGCTGCGGTCAATCTTAAAC 660
DB 601 AAGGGAATATTAGATATGTTACTGCACTTAAAGTCAAGCTGCGGTCAATCTTAAAC 660

```

Db 4796 AAGGAAATATTACAGATATGCTTACCTCAGTTCAAAAGTGACGCTCGCTGCATCTTAAAC 4855
Qy 661 TTGCGTCCAACTGTGTGGGGGACACATATATTTGGAGAAAATTCGTGTGATATGCTTTTAT 720
Db 4856 TTGCGTCCAACTGTGTGGGGGACACATATATTTGGAGAAAATTCGTGTGATATGCTTTTAT 4915
Qy 721 GATGATATGATCTAACAAGCAGCTCTTTGGAGATTAAGATTTGAGATTAACATCTTAA 780
Db 4916 GATGATATGATCTAACAAGCAGCTCTTTGGAGATTAAGATTTGAGATTAACATCTTAA 4975
Qy 791 TGTGATGGAATTTTATCTAAGAAAATTAATGATGACACCAAGAAATTCATATCT 840
Db 4976 TGTGATGGAATTTTATCTAAGAAAATTAATGATGACACCAAGAAATTCATATCT 5035
Qy 841 TTGTACATCTCTTGGCGGGTAAAGTTAACTCCACAAATGAGACGTCAATTAATAT 900
Db 5036 TTGTACATCTCTTGGCGGGTAAAGTTAACTCCACAAATGAGACGTCAATTAATAT 5095
Qy 901 GGTGACGAGCTTCTCTGGAACCAACCTGGAATTAATTAACAGCTGTCAACCTGCAAGA 960
Db 5096 GGTGACGAGCTTCTCTGGAACCAACCTGGAATTAATTAACAGCTGTCAACCTGCAAGA 5155
Qy 961 ATCACTGTTCCGCTGTGTGTGCTGACGCTTGAATTCGATTCGAAAGTGGAAAT 1020
Db 5156 ATCACTGTTCCGCTGTGTGTGCTGACGCTTGAATTCGATTCGAAAGTGGAAAT 5215
Qy 1021 CCGGAGCTGACCAATATATATGATTAATTAATTTACTTCCACCAAGTAGTCAACA 1080
Db 5216 CCGGAGCTGACCAATATATGATTAATTAATTTACTTCCACCAAGTAGTCAACA 5275
Qy 1081 CTCTAG 1086
Db 5276 CTCTAG 5281

RESULT 3
AX352351 7239 bp DNA linear PAT 06-FEB-2002
LOCUS Sequence 27 from Patent W00181582.
DEFINITION AX352351
ACCESSION AX352351 GI:18617634
VERSION
KEYWORDS
SOURCE
ORGANISM
Escherichia coli
Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 Alboum, Z., Levine, M.M. and Barry, E.M.
Isolation and characterization of the csa operon (etec-csa pl1)
and methods of using same
Patent: WO 0181582-A 27 01-NOV-2001;
University of Maryland, Baltimore (US)
LOCATION/Qualifiers
1. 7239
/organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"
283. .999
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAD22859.1"
/db_xref="GI:18617635"
/translation="MHKLFCLSLITPFPVANFMIVPISKDKNGNSSEILIRYSKS
KEIQYIKYTKIKINPTEBEHEVDMMDGFPVTPQKVLIPAGSSSILRQFRI
KKEIYRVFAVRKPSKENVYDNKLTETSEVNIYALIRSLPSEINISNTSRNA
RKIIYVNGNVRAGVNDYKPCSKSNIDSCVKTHKNNTYPERKSFILVNNNSYV
IKLNHDEKXQGLQKVP"
1028. .1531
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAD22860.1"

CDS

Db 4796 AAGGAAATATTACAGATATGCTTACCTCAGTTCAAAAGTGACGCTCGCTGCATCTTAAAC 4855
Qy 661 TTGCGTCCAACTGTGTGGGGGACACATATATTTGGAGAAAATTCGTGTGATATGCTTTTAT 720
Db 4856 TTGCGTCCAACTGTGTGGGGGACACATATATTTGGAGAAAATTCGTGTGATATGCTTTTAT 4915
Qy 721 GATGATATGATCTAACAAGCAGCTCTTTGGAGATTAAGATTTGAGATTAACATCTTAA 780
Db 4916 GATGATATGATCTAACAAGCAGCTCTTTGGAGATTAAGATTTGAGATTAACATCTTAA 4975
Qy 791 TGTGATGGAATTTTATCTAAGAAAATTAATGATGACACCAAGAAATTCATATCT 840
Db 4976 TGTGATGGAATTTTATCTAAGAAAATTAATGATGACACCAAGAAATTCATATCT 5035
Qy 841 TTGTACATCTCTTGGCGGGTAAAGTTAACTCCACAAATGAGACGTCAATTAATAT 900
Db 5036 TTGTACATCTCTTGGCGGGTAAAGTTAACTCCACAAATGAGACGTCAATTAATAT 5095
Qy 901 GGTGACGAGCTTCTCTGGAACCAACCTGGAATTAATTAACAGCTGTCAACCTGCAAGA 960
Db 5096 GGTGACGAGCTTCTCTGGAACCAACCTGGAATTAATTAACAGCTGTCAACCTGCAAGA 5155
Qy 961 ATCACTGTTCCGCTGTGTGTGCTGACGCTTGAATTCGATTCGAAAGTGGAAAT 1020
Db 5156 ATCACTGTTCCGCTGTGTGTGCTGACGCTTGAATTCGATTCGAAAGTGGAAAT 5215
Qy 1021 CCGGAGCTGACCAATATATGATTAATTAATTTACTTCCACCAAGTAGTCAACA 1080
Db 5216 CCGGAGCTGACCAATATATGATTAATTAATTTACTTCCACCAAGTAGTCAACA 5275
Qy 1081 CTCTAG 1086
Db 5276 CTCTAG 5281

BASE COUNT 2436 a 1181 c 1345 g 227 t
ORIGIN

Query Match 100.0%; Score 1086; DB 6; Length 7239;
Best Local Similarity 100.0%; Pred. No. 3.3e-223;
Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAATAGATTTATTTATTTTACATGTTTCTTCAGTACTTTTACATTTGCT 60
|||||
4196 ATGAATAGATTTATTTATTTTACATGTTTCTTCAGTACTTTTACATTTGCT 4255
|||||
61 GTATCGCAGATTAATATCCCGGAGATGAAGCATTAATATTTTGGCCCGCGTAC 120
|||||
4256 GTATCGCAGATTAATATCCCGGAGATGAAGCATTAATATTTTGGCCCGCGTAC 4315
|||||
121 AGGAACGAATCTTCCCAACATATATTAATTAACATTTTACAGCTAAGTGA 180
|||||
4316 AGGAACGAATCTTCCCAACATATATTAATTAACATTTTACAGCTAAGTGA 4375
|||||
181 AGTCACTCTGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
|||||
4376 AGTCACTCTGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4435
|||||
241 GGAGCAGTCCCAACAGTGAAGATCTGAGCTTATCGGTCAGCGGTGAACAAATATA 300
|||||
4436 GGAGCAGTCCCAACAGTGAAGATCTGAGCTTATCGGTCAGCGGTGAACAAATATA 4495
|||||
301 ACATTTACATTTACGGAAGAAAGATTATTAAGAGAGCTACAAATTAAGGCTAT 360
|||||
4496 ACATTTACATTTACGGAAGAAAGATTATTAAGAGAGCTACAAATTAAGGCTAT 4555
|||||
361 AAACATTTATGTTCAAAAGTGAATGATGATGATGATGATGATGATGATGATGAT 420
|||||

CDS

Db	4556	AAACATTTTCTTCAAAAAGTTTAACTGCCATCCGCCCTAACACTTAATCAGCTCAT	4615	gene	283. .999	/gene="csaa"
Qy	421	TTTAACGTGAATAAAAACGCGGCTTCAGGTGCAAGTTTATATATATCTCTGCTGCG	480	CDS	283. .999	/gene="csaa"
Db	4616	TTTAACGTGAATAAAAACGCGGCTTCAGGTGCAAGTTTATATATATCTCTGCTGCG	4675			/codon_start=-1
Qy	481	GAACATAAAAAATTTGCTTTTGGTGGATCTGGAGTCTACTCTGAGTTAAAGATAA	540			/transl_table=-1
Db	4676	GAACATAAAAAATTTGCTTTTGGTGGATCTGGAGTCTACTCTGAGTTAAAGATAA	4735			/protein_id="CAD90914.1"
Qy	541	AACAGATATAGTGAACCTATGAACTTACTATATATATATATATATATATAT	600			/db_xref="GI:30524216"
Db	4736	AACAGATATAGTGAACCTATGAACTTACTATATATATATATATATATATAT	4795			/translation="MKKLCGLSLITLTPVANANMTPYISDKLNKNGSELIRVYSKS
Qy	601	AAGGAAATATTCAGATATGTTACTCTCAGTTCAAAGTGAAGCTGCGCTCATCTTAC	660	gene	1028. .1531	KEIOYIKIYTKRIINPGTTEHEHVDMPNMDGFEVTPPOKVIILPAGGSKSIRTORIP
Db	4796	AAGGAAATATTCAGATATGTTACTCTCAGTTCAAAGTGAAGCTGCGCTCATCTTAC	4855	CDS	1028. .1531	KKEIYIRYFEAVKPKDSKENYIDNKKLTLSVNIITAYALISLSEQISLISRNA
Qy	661	TTGCGTCCAACTGGTGGGGGCAACATATATTTGAAGAATTCGTTGATATGCTTTTAT	720			RNNITLYNNGNRAVGKDIYFCKSSNIDSCVKKTHNNKIYPEKSFDTLLVNNFSYVF
Db	4856	TTGCGTCCAACTGGTGGGGGCAACATATATTTGAAGAATTCGTTGATATGCTTTTAT	4915			IKLNHDEIEKGLIOLKVP"
Qy	721	GATGATATAGTACTAACAGCAGCTCTTGGAGATTAAGATTTTCAAGATTAACATCTAA	780			1028. .1531
Db	4916	GATGATATAGTACTAACAGCAGCTCTTGGAGATTAAGATTTTCAAGATTAACATCTAA	4975			/gene="csab"
Qy	781	TCTGATGGGAATTTTATCTAAGAAAAATATATGATGACACCAAGAAATTCATATCT	840	gene	1589. .4192	/codon_start=-1
Db	4976	TCTGATGGGAATTTTATCTAAGAAAAATATATGATGACACCAAGAAATTCATATCT	5035	CDS	1589. .4192	/transl_table=-1
Qy	841	TTGTCACTTCTTGGCGGGTAAAGTTTAACTCCAACAATGGAACGTCAATTAATAT	900			/protein_id="CAD90915.1"
Db	5036	TTGTCACTTCTTGGCGGGTAAAGTTTAACTCCAACAATGGAACGTCAATTAATAT	5095			/db_xref="GI:30524217"
Qy	901	GCTGACGAGCTCTCTGGAACAACTGGATATAGATATACAGCTGTCACCATGCCAGAA	960			/translation="MKIKTIGAMALITTFVAMSASAVEKNITTYASVDPTIDILAD
Db	5096	GCTGACGAGCTCTCTGGAACAACTGGATATAGATATACAGCTGTCACCATGCCAGAA	5155			GSSLPTAVELITSPASPRENKIKATKHTINYNKNAVKLVNDPKLTINVDSTOLP
Qy	961	ATCAGTGTCCGCTGTGTGTGTGGCTGAGCTTTGCAATTTGATGACAAAGTGGAAAT	1020			ITVSYGKTLSTADVTPEPAELNFTSGVTGVSSQDLVIGATTAQAPTAGNYSGVYS
Db	5156	ATCAGTGTCCGCTGTGTGTGTGGCTGAGCTTTGCAATTTGATGACAAAGTGGAAAT	5215			ILMTLAS"
Qy	1021	CCCGAGGCTGGCAATATATGCGTAATATTAATGTTACTTTACACACCAAGTGTCAACA	1080	gene	1589. .4192	/gene="csac"
Db	5216	CCCGAGGCTGGCAATATATGCGTAATATTAATGTTACTTTACACACCAAGTGTCAACA	5275	CDS	1589. .4192	/codon_start=-1
Qy	1081	CTCTAG 1086				/transl_table=-1
Db	5276	CTCTAG 5281				/protein_id="CAD90916.1"
LOCUS	AX741418	7239 bp	DNA	linear	PAT 10-MAY-2003	/db_xref="GI:30524219"
DEFINITION	Sequence 8 from Patent WO03022306.					/translation="MNKILFITLFPSSVLTFFVYASADIRPDDESTINTFGRDRNES
ACCESSION	AX741418					SPKRNLTNNHTIAYSESHTLIDRMATFLCSSHNTLNGACPISENBSSSVSGETNITL
VERSION	AX741418.1	GI:30524215				OPTERKSLIKRELQILKGYKQLLFKSYNCVCSGLTUNSAHNCNKMAASLVLIPAG
KEYWORDS						ELKLNLPFGGTYIMPATILKRVKRRYSSEYGYTINITIKLIDKNGINIOAGLPSDARVD
SOURCE						LNMLRPFGGTYIGRNSVDVCMCFYDGYSTNSSLSEIRIODNNPKSDGFEYLRKINDPKE
ORGANISM	Escherichia coli					IAYTSLILAGKSLIMPNTGNTSLINDAASLELTNNMNRITAVTNPBELSVPLVCPRQLQ
REFERENCE	1	Turner, A.K., Greenwood, J., Stephens, J.C., Beavis, J.C. and				DAKVENPEAGCYMGNGNTNVTFFPSSQL"
AUTHORS	1	Darley, M.J.				6346. .6649
TITLE	Bacterial Vaccine					/note="unnamed protein product; ISI sequence"
JOURNAL	Patent: WO 03022306-A 8 20-MAR-2003;					/codon_start=-1
FEATURES	Acambis Research Limited (GB)					/transl_table=-1
source	1. 7239					/protein_id="CAD90918.1"
	Location/Qualifiers					/db_xref="GI:30524220"
	/organism="Escherichia coli"					/translation="MNGSPHYGRMPQHDFTSLKLRLPOSTVTSRIQPGSDVIVCAEMD
	/mol_type="genomic DNA"					EQMGYVGAASRQRLMEYAYDRLKRTVAVAFGERMATLIGRLMSLSLSPEDVIVMTDG
	/db_xref="taxon:562"					WPLYESRLKGLHVISKRYTORIERHNLTROHLARLGRKSLSPFSKSVELHDKVIGHY


```

Db      2883 AGCTATATGACAGATTTGTTTTTATGCTACATCTCTCCAAATCCGGTAAATGCTGCT 2942
QY      247 TGTCCAAACGAGTGAATTCCTAGACTGATCGGTGACGGTGAACAAATATACATTA 306
Db      2943 TCCCAACCATTTGGAACATCTGGAGTCAAT-----ACGGTACTCAACCAATACCTTG 2996
QY      307 CAATTACGGAAGAAAGATTATTAAGAGAGCTCAATTAAGAGCTATTAACA 366
Db      2997 CAGTTTACGAAAGAAAGTCTGATTAAGAAATATTAATCTGACGATTAAGAA 3056
QY      367 TTAATGTTCAAAAGTTTAACTGACCGCATCCGCTTAACACTT-----AACTCAGTCAAT 420
Db      3057 CCAATATGGAGAAATCAGATTGCGACTTTAGCAATCTAATGGTGTGAATTCGAAGTCT 3116
QY      421 TTTAACTGTAAATAAACG---CGGCTGAGGTGCAAGTTTAAATTAATATATCTCTGCT 477
Db      3117 TGGAGCTGTGGGGCTCAGGAATGCTAACGGAAACCTTAATCTGATATATCCCTGA 3176
QY      478 GCGCACTAAAAAATTTGCTTTGCTGATCTGGGATGCTACTGTAAGTAAAGTA 537
Db      3177 GGAGAAATCAACAATTTGCTTTGGAGGATATGGAGCAACTCTGATCTTACGCTTA 3236
QY      538 AAAAG-----ACGATATAGTAGACCTTAATGAACTTAACATTAATATCACTATT 588
Db      3237 TCAAGATATGGGAAAGTCACTAGACCCCACTTACGGCAATTAATACCTTAATATTCGCTT 3296
QY      589 AAATTAAGTAAAGGAAATATTCAGATATGCTTACCTCACTCAAAAGTACGCTCC 648
Db      3297 GATTTAACTGATTAAGGTAATATTCAGTATGCTTCCAGGCTTACAGCAACCGGCT 3356
QY      649 GTGCAATCTAATCTGCTCCAACTGCTGGGGGACACATATTTGGAAGAAATTCGTGAT 708
Db      3357 GTAGACCTGAATCTGGCCCTATCGGTAAATTAATTAATAGTGAATTCACATCAGCAG 3416
QY      709 ATGCTCTTTATGATGATATAGTACTAACACAGCTTTGGAGATTAAGATTTAGAT 768
Db      3417 ATGCTTTCTATGATGATATAGTAAACAAACATGATGATATTAAGTTCCAGGAT 3476
QY      769 AACAATCTAAATCTGATGAGGAAATTTATCTAAGGAAATTAATGATGACACCAAGAA 828
Db      3477 GATTAATCTCAACCAATTCATCTCAATTAATTAATCTTTAATTAAGATTA---GGGGGCACTGA 3533
QY      829 ATTGATATATCTTTGCTGCTCTGCGGGGTAAGTTAACTCAACCAATTAAGTAAG 888
Db      3534 TTACATATGCTGCTTCACTGCTTATGGGAGAAATTTTATTCACATGATGCTCAA 3593
QY      889 TCATTAATATTTGCTGACAGCTCTCTGGAACCAATCTGGAATGAATTAACAGTCTC 948
Db      3594 TCATTTACTATCAATGACAGTATGCTGCTGGAACCAATCTGGAATGAGTAAACGCAATT 3653
QY      949 ACCATGCCAGAAATCAGTGTCCGGTGTGTTGGCTGAGACGTTTGCATTTGATGCA 1008
Db      3654 GGTATCCCGGAAGTTAATGTTCCAGATTAATGCTGCGCAGCAAGATTTGATTAATGCT 3713
QY      1009 AAAGTGAAGAAATCCGAGGCTGACATTAATGCGTAATTAATTAATTTACTTTACACCA 1068
Db      3714 GATGTAAAGCTCCCGATGACAGAGTATTCAGCAGATATATTAACATTTACACCC 3773
QY      1069 AGTAGTCAAA 1078
Db      3774 AGTGTGAAA 3783

```

```

RESULT 7
AX741413
LOCUS      AX741413      5336 bp      DNA      linear      PART 10-MAY-2003
DEFINITION Sequence 3 from Patent WO03022306.
ACCESSION AX741413
VERSION    AX741413.1  GI:30524203
KEYWORDS
SOURCE
ORGANISM   Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

```

```

REFERENCE
1
AUTHORS    Turner,A.K., Greenwood,J., Stephens,J.C., Beavis,J.C. and
            Darstey,M.J.
TITLE       Bacterial vaccine
JOURNAL     Patent: WO 03022306-A 3 20-MAR-2003;
            Acambis Research Limited (GB)
FEATURES
source
            /organism="Escherichia coli"
            /mol_type="genomic DNA"
            /db_xref="taxon:562"
            85..2703
            /note="unnamed protein product; cooc"
            /codon_start=1
            /transl_table=11
            /protein_id="CAD90906.1"
            /db_xref="GI:30524204"
            /translation="MIGKSSRVAVIVLSYIGSSGFAKRYNLVDIPESFRLMGEOD
            ELLEVRVLSGVHRIKSTPTTVAEESPDLNLDIEIKRGKREADLRVLMRSPQRNG
            NMSCGYTGONNCNTIKTNTVAIVDDVENVNLPIGNEFLASGENSDYXOPSKNTK
            KAFISQITLSDTGNVEMLSVIGSGISGITNSYAILGMAANYNRYKSYANTNEDSIN
            SLVFRHDEFKNEFYOLGRIDRSDLSQSSGGMNPDLIPDOIYGRACGTQSYIKNTG
            KSYASPTTLMHRSREAYRNGQLGWTLDAGISELTERLPBNTDLKLTFOE
            QVREIIVFNKSGSITGTHNDVFOADIIINDGRIVEKKNRKSALNSGLRPLT
            RNLAVOLGAVIDNKNRYETGLIMNSGIDGSLNSKFTFLFDGDDHGNQYNSYTDG
            SLSPFVNDKRVDCDGNMNGSGCESYASLSIPVKMNSTLAYSNYSYSTRYD
            AVSEVYVYVYKGRKMQLTASTVYRMGDYNIIMPTIGVYNSKOROMADKGYLSLT
            TVDQGSLSNAGYSYNSRGNTSNDAPVEGHLVDQNTVRELSARVSGNRYTEGG
            VSGRINNRGDIAGTLNVKNRKSHDTHLSLAGYSSPALTDGITYNGASGLTNL
            SGGIYRVKNEDESELNAYKGSYGHYSLSGSDSLFIPPALMQSLTIEENTNSKN
            IDVLAPTKTFEFLPESVYPIDVASNVSTYVGRVAKGRPLSGAYILMAQNTYLDN
            NGFSESESENKELEFLDKRTIYCSLDRSBRNGIAFVEGVAACNSTIKELLPKLV
            TNSRIHDLIYQDTE"
            2700..3791
            /note="unnamed protein product; cood"
            /codon_start=1
            /transl_table=11
            /protein_id="CAD90907.1"
            /db_xref="GI:30524205"
            /translation="MKKIPFLSIFSAVAVSAGRYPETVGNLTKSPQAPRLDSVOS
            PLYNFTNNAVYSLSLSLXDRIVFLCTSSSPVNVACPTCTSGVOGTITTTQFT
            EKRSLIKRININAGNKPRLWENQSDSFSLNLSKNSMCGHANGNLWLTLPAG
            ELNKLPEFGIWEATLILRLSRGSEVSTHYVNTVITVDLKDKNINQWLPFGFSNP
            RVDLNLRLPIGNKTVSGSNSLDMCFYDGTSTNSDSWIKIKODPNPNSSSYNLKIGCT
            EKLPLVASLIMEKILYPVNGOSFTINDSVIETNMNRVTAVAMEVNVPIVCPARL
            LNDVADNADAOYSGQIYITFTPEVNL"
            3791..4201
BASE COUNT  1553 a 1021 c 1206 g 1556 t
ORIGIN
Query Match      28.6%; Score 310.8; DB 6; Length 5336;
Best local Similarity 58.8%; Pred. No. 1.1e-56;
Matches 641; Conservative 0; Mismatches 422; Indels 27; Gaps 5;

```

QY	307	CAATTTACGAAAAAGAGTTTATATAAAGAGAGCTACAAATTAAGGCTATAACAA	366
Db	2997	CAGTTTACAGAAAAAGAGAGTCTGATATAAAGAAATATATATCTTCCGAGTAATAGAAA	3056
QY	367	TTATTTGTCAAAAGTGTTAACTGCCATCCGGCTAACACTT-----AACTCAGCTCAT	420
Db	3057	CCAAATATGGGGAAATCAGAGTTGGAGACTTTAGCAATCTAAAGTGTTGAATTGCAAGTCT	3116
QY	421	TTTAACTGATATAAAAAACG---CCGGCTCAGGTGCAAGTTTATATATATATATCCGCT	477
Db	3117	TGAGAGCTGTGGGGCTCAGAGAAATGCTAACGGAAACACTTCTAATCTGTATATCCCTCA	3176
QY	478	GGCGAACTAAAAAATTTGCTTTGGTGATATCTGGAGTGTCTACTGTGAATTAGACTA	537
Db	3177	GGAGAAATCACAAAATTTGCTTTGGAGGGATATGGGAGCAACTGTATCTTACGGCTTA	3236
QY	538	AAAAG-----ACGATATAGTAGAGACATATGAACTTACACTATATAATCTACTTT	588
Db	3237	TCAAGATATGGCGAAGTCATGAGCAACCCATTACGGCAATTTATCCGTAATAATTAACGGTT	3296
QY	589	AAATTTAATCGATTAAGGGAATATTCAGATATGTATGTTACCTCAGTTCAAAAGAGACGCTGC	648
Db	3297	GATTTAACTGATATAAGATATATTCAGGTATGCGTTCCAGGGTTTCAACAGAACCCGGT	3356
QY	649	GTCGATCTTAACTTGCCTCCCAACTGCTGGGGGACATATATTTGGAAGAAATTTGTTGAT	708
Db	3357	GTAAACCCGAAATCTGCCCTATCCGTATTAATATATATAGTGTAGTAATTAACCTGCAC	3416
QY	709	ATGTGCTTTATGATGATATAGTATCTAACAGACAGCTCTTTGGAGATTAAGATTCAGGAT	768
Db	3417	ATGTGTTCTATGATGATATAGTATAGTATCAACACAGTATAGCATGTTATTAAGTCCAGAT	3476
QY	769	AACAATCTTAATCTGATGGGAAATTTATCTAAGGAAATTAATGATGACACCAGAAAG	828
Db	3477	GATATATCTTCAACCAATTCATCTGAAATATATCTTTATAAGATA---GGGGGCACTGAAAA	3533
QY	829	ATTCGATATATTTGTGTGCTCTCTTGGGGGTTAAAGTTTAACCTCCACAAATGAGAACG	888
Db	3534	TTACCAATATGCTGTTTCACTGCTTATGAGGAGAAAAAATATTTATCCAGTAATGCTCAA	3593
QY	889	TCATTAATATATTTGCTGACGACAGCTTCTCTGGAACCAAACTGGAATAGATTAACAGCTGC	948
Db	3594	TCATTTACTATCATCATACAGATAGTATGACTCGAAMCAAACTGGAATCGAGTAACGCAATT	3653
QY	949	ACCATGCCAAGAAATCAGTGTCCGCTGTGCTGTGCTGGACGTTTGCAATTGGATGCA	1008
Db	3654	GCTATGCCGGAAGTAAATGTTCCAGTATTAATGCTGGCCAGCAAAATTCCTATTAATAGCT	3713
QY	1009	AAATGGAATATCCCGAGAGCTGACCAATATATATGATTAATATGTTACTTCACTCAACCA	1068
Db	3714	GATTAATATGCTCCCGATCGACGACAGATATCAGAGACAGATATATATTAACATTACACCC	3773
QY	1069	AGTAGTCMAA 1078	
Db	3774	AGTGTGMAA 3783	
RESULT 8			
LOCUS	AY216491	6956 bp	DNA linear BCT 12-MAR-2003
DEFINITION	Escherichia coli strain DS37-4 plasmid CS17 gene locus, complete		
ACCESSION	AY216491		
VERSION	AY216491.1	GI:28932774	
KEYWORDS			
SOURCE	Escherichia coli		
ORGANISM	Escherichia coli		
REFERENCE	Bacteriae; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
AUTHORS	Enterobacteriaceae; Escherichia.		
TITLE	1 (bases 1 to 6956)		
	Biediger W. and Wolf M.K.		
	CS17 of Enterotoxigenic Escherichia coli		

JOURNAL Unpublished
2 (bases 1 to 6956)
REFERENCE Biediger, W. and Wolf, M.K.
AUTHORS Direct Submission
TITLE Submitted (09-JUN-2003) Enteric Infections, Walter Reed Army
JOURNAL Institute of Research, 503 Robert Grant Ave, Silver Spring, MD
20910-7500, USA

FEATURES
source location/Qualifiers
1. 6956
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="DS37-4"
/db_xref="taxon:562"
/plasmid="unnamed"
/note="naturally-occurring plasmid"
166. 879
/gene="csbB"
166. 879
/gene="csdB"
/note="CS17 fimbrial accessory protein"
/codon_start=1
/transl_table=11
/product="CsbB"
/protein_id="AA060095.1"
/db_xref="GI:28932775"
/translation="MKRLFLSLMLPEPTVKNFMYPISEIKGASSELLIRYSKND
TOYIKYAKKRVGPTGRDEXEIDAPNMEGALVYTPAVILPSSGSKSVRLTQLAAPT
EEYRYAFVESVKEPAQENSNGTKLKDLSVNYIALILPLPINKTASVSISPEN
NLIKNTGNVRLGIMVYFCFKSSINDCYKSKKNVYGYLLDFWIYNDGYSHLFL
DTEDSSKNRDLNLYPIIT"
933. 1439
/gene="csbA"
933. 1439
/gene="csbA"
/note="CS17 fimbrial subunit; structural subunit of CS17
plus"
/codon_start=1
/transl_table=11
/product="CsbA"
/protein_id="AA060096.1"
/db_xref="GI:28932776"
/translation="MKRLKIGAMAAATLEPATMAASAVEKITVYRASVDPEKLLDAD
GSLPSIALTYSSANFEVYSILNTAIIHNDKSGVYKLSAPVLSNIMKNSOIP
MKYTLGGKTLNTIDTERTYDTLNFGTSGVENSVTQDLTHADTQGTAPAGNQTGII
SLIMTKRT"
1506. 4124
/gene="csbC"
1506. 4124
/gene="csbC"
/note="putative fimbrial usher"
/codon_start=1
/transl_table=11
/product="CsbC"
/protein_id="AA060097.1"
/db_xref="GI:28932777"
/translation="WYSKNSKLSIYLEFLSSPLGPAQKNKTTINPESRDLMGEDD
GLEVEYFGOSLGVRHKLKLEPTVAFESPSPSLDKLEINKERADRLVIMRGSPQRNG
MNSCGYTGQNNCNIKTNINVAIVDDVENLVNLFIGNERLASGENDSDIYQPSKNTK
KAIHSDTILNLDPTGNVENLISIVGSLGTTDSVAIIQGAANYNRYKSYNVEQSIN
SLYFRDPEKNEYQDGRIDRDLSSQSGENFEDLLPVDIDYMGAGTQSYIKKTG
KSYASVPTIMLITHFSSVEAYRNGQLLGWYLDAGISELDERLPDGNYDLKLFPOE
QLVREELIYFPNKSIGSDTHMDVFOVQAGIINDNRYVEKQKHKSAINSLGRLPLT
RNAVLQGLGAVIDNKNVYETGILMNSGFLDGLSINSKPTFLPGDTHGCAONVSYTGF
NLSFYNDKRVNDCGDYNNMGSCYSESASLSITPYKGNSTLAIASNTISYTRID
ANSEYPISTYKGRTRKWLDTASTVVRMGDYNIMPTIGYNSQKQKQADGGLSLTL
TRVDGKSLINAGSYNYSRGNYSNDAFEGRVSDTNSYREL.SARVASGNRYTREG
VSGRIINRGEDLNGTLNKNRSKSHDTSHLITAGYSSFALTGDIYWGGSAGLNLN
SGCIVRYKSNDESDLELVNKGSSYGNYSLSGNSDLFIPVALQASITILEKNTSKN
IDVLAPTKNTFFMLPGSVYPIDVASVSPFYVGRGVDRPLSGAIVLINAQIVIDE
NGGFSFESSENEKEFLDKDTIYSCSLDRSEARNQIVPGEVACNSTIKELLPEKIV
TNSRHIDLAAYNDTE"
4121. 5212
/gene="csbD"

QY	538	AAAGACGATTTAGTGGACCTTATGGAACTGACCTTAAATATACATTAATTAATCT	597
Db	4997	AAAAGCATTCACGCTATATATCATGATGTTCTTCAAAAGTTATATACAGTTGATTTGACA	5056
QY	598	GATTAAGGAAATATTTACAGATATATGGTTACCTCAGTTCCAAAAGTGAAGCGTCCGTCGATCTT	657
Db	5057	GACAAAGAAATATATTCAGGTCGTGGACACCAAAAGTTTCATATAGCGATTCCTAGAAATTTGATCTG	5116
QY	658	AACCTGGCTCACTGGTGGGGGACATATTTGGAAAGAAATTTCTGTTGATATGTCCTT	717
Db	5117	AATTTAGCTCTCGAAGGTAAATGTTAAATATTTCTGCTAGTAAACGTGCTGAGATGTCCTC	5176
QY	718	TATGATGATATATAGTAAACAGCAGCGCTTTGGAGATTAAGATTTGAGATTAACGATTAACATCCCT	777
Db	5177	TATGATGATGATATAGTAAACAGTATGATTAAGTAAATGAGCTTTCAGATGATCACA	5236
QY	778	AAATCTGATGGGAAATTTTATCTTAAGGAAATAAATGATGACACCAAGAAATTTGCATAT	837
Db	5237	ACAGGAATATATGAAATATTAATCTTAATAAAACTGGAGGCCATTTAAAAAATTTGCCATAT	5296
QY	838	ACTTTGTCATCTTCTCTGGCGGGTAAAAAGTTTAACCTCCACAAATGGAAGCTCATTAAT	897
Db	5297	AAACTTTCTCTCTTTTAAAGAGGACGAGAGTTTATCCAAATTAATGGAGAGGCTTTTACT	5356
QY	898	ATTGCTGACCCAGCTTCTCTGGAACAAACATGGAATGAATTAACAGCTGCACCATGGCA	957
Db	5357	ATTATATATATCTGCTATATGTTTATTAACCTGGAATGATTAAGTCTGTTCTTTACCA	5416
QY	958	GAAATCAGTGTCCGGTGTGTGTGTGGCCCTGGACGTTTGCATTTGATGACAAAAGTGGAA	1017
Db	5417	CAGATTGTAGTATTCAGTACTATGCTGGCCAGCAAACTTGACATTTATGTCAGAGCTTAAT	5476
QY	1018	AATCCGACAGCTGGACATATATATGGTATATTTATGTTACTTTACACCAAGTAGTCAA	1077
Db	5477	AATCCAAAGCGGGTGATTTACAGGAATACTTTAAACGTAACATTAATTAATCTCAATGATTTCA	5538
QY	1078	ACACTCTA 1085	
Db	5537	AGTCTGTA 5544	
RESULT 10			
AX741415			
LOCUS	AX741415	5798 bp	DNA
DEFINITION	Sequence 5 from Patent WO03022306.	linear	PAT 10-MAY-2003
ACCESSION	AX741415		
VERSION	AX741415.1	GI:30524208	
KEYWORDS			
SOURCE			
ORGANISM	Escherichia coli		
	Escherichia coli		
	Bacteria: Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
	Enterobacteriaceae; Escherichia.		
REFERENCE			
AUTHORS	1 Turner,A.K., Greenwood,J., Stephens,J.C., Beavys,J.C. and		
	Darsley,M.J.		
TITLE	Bacterial vaccine		
JOURNAL	Patent: WO 03022306-A 5 20-MAR-2003;		
FEATURES	Acambis Research Limited (GB)		
source	Location/Qualifiers		
	1..5798		
	/organism="Escherichia coli"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:562"		
	499..1215		
	/gene="cotB"		
	/gene="cotB"		
	499..1215		
	/gene="cotB"		
	/transl_table=1		
	/codon_start=1		
	/protein_id="CAD90909.1"		
	/db_xref="GI:30524209"		
	/translation="MKILLVPILEFNVFAASANEMVYPISKIDSGSEETIKVFSKSK		

Query Match	Similarity	Score	DB	Length
Best Local	58.5%	Pred. No. 4e-54;		
Matches	566;	Conservative	0;	Mismatches 387; Indels 15; Gaps 2;
Db	127	GAATCTTCCGCCAACAATAATATATTAATAACATATTTACAGATACAGTGAAGCAT	186	
Db	4583	GCATCGTATCTCGTCAATATATATTCATGAACATGTGTGGTTCAATAAAGATCAC	4642	
Qy	187	ACTCTGATGATAGAGATGACTTTTATATGTTTGTCTTCTTCACAAATTAACCTTAAGGAGCA	246	
Db	4643	TCTCTTTTTCACAGAGAGACGTTTATATGATGATCATCAACAAGATGATCAAAAGTGCA	4702	
Qy	247	TGTCCAAACAGTGAATCTCTAGCATTCATCGGTCAAGCGGTGAACAATAATATTAACATTA	306	
Db	4703	TGTCCGACAGGAGAAACTCCA-----AATCTCTCAAGGAGGAGACTAAATATTAACCTA	4756	

OY		307	CATTTCAGGAAAAAAGATTAAATTAATAAAAACAGAGCTACAAATTAAGAAGCGTATTAACA	366
Db		4757	AATTTACTGAAAAAAGTAGTGCCGCAAAAAAACATTAACTTAAAAAGATATTAAGGA	4811
OY		367	TATTTGTTCAAAAAGTTTAACATGC-----CCATCOCGGCTTAACCTTAACACAGCT	417
Db		4817	TTTTTATATATGAAATCAGATAGATGATCATATTAATGTGCAATTAATAATGAAATCTCAATTCAT	4876
OY		418	CATTTTAACCTGTAATATAAAAACGCCGCTCAGGTCGAACTTATATTTATTTATTTCTGCT	477
Db		4877	ACTGTTAAATGTGTAGGTTCAATTCACAAGAGAGAGATGATTTTCACCTTTATATATCCACAA	4936
OY		478	GCGCAACTAAAAAAATTTGCTTTTGGTGGTATCTGGATATCTGACCTCAAGTTAAGAGA	537
Db		4937	GGTGAATATGATNGGGCTCTCACTAGGAGTATATGAGAGCCACACCTAGAGTTACGAGTC	4996
OY		538	AAAAGACGATATATGTAGACCTATGGAACCTTACACTATTAATATACATATTAATTAAT	597
Db		4997	AAAAGGCAATTTACGACTATATATCATGTGTACTTACAAAGTTAATATCACAGTTGATTTGCA	5058
OY		598	GATTAAGGAAATATTCAGATATGTTTACTCTGATTCATTAAGGAGACGTCGGGTGATCT	657
Db		5057	GCAAAAGGAATATTCAGAGTCTTGACACCACCAAAGTTTCATPAGGATCTGAAATTAATGTG	5118
OY		658	AACTTGCCTCAACATGCTGGGGGACACATATATTTGGAAAGAAATTCGTGATATGTCTTT	717
Db		5117	AATTTACGCTCGAAGGTAAATGTTAAATATCTGTGATTAAGCTGCTTGAATGTGTCTC	5176
OY		718	TATGATGATATATAGTACTAACAGCAGCTCTTTGGAGATTAAGATTTACAGATTAACATCT	777
Db		5177	TATGATGCTATATAGTACACATATGTCMAAGTATAGAAATGAGGTTTACAGATGACTCACAA	5238
OY		778	AAATCTGATGGGAATTTTATCTTAAGAAAAATTAATGATGACACCAAGAAATATGCATAT	837
Db		5237	ACAGGAATTAATGATATATATCTTTAAAAACTGAGAGCCATTTAAAAAATTTGCCATAT	5296
OY		838	ACTTTGTCACTTCTCTGCGGGGTTAAAGTTTAACCTGCAACAAATGAGAACTCATTAAT	897
Db		5297	AAACTTCTCTCTTTTGAAGAGACGAGAAGTTTATCCAAATTAATGAGAGGCTTTTACT	5356
OY		898	ATGTGTCAGCGACTTCTCTGGAACCAACATGGAATTAATTAACAGCTGTCAACATGCCA	957
Db		5357	ATTAATGATPACTGCTCATATGTTTAAACATGGAATTCGATTAATGCTGTATCTTTACCA	5418
OY		958	GAAATCAGTGTCCGCTGTGTGTGTGGCCTTGACCTTTGCCAATTTGGATGCAAAAAGTGGAA	1017
Db		5417	CAGATTAGTATTTCCAGTACTATATGCTGGCCAGGAACCTTGACATTTATGTCCAGAGCTAAT	5478
OY		1018	AATCCCGAGCCTGCGCAATATATGTGGTAAATTAATGTTACTTTCCACACCAATAGTCAA	1077
Db		5477	AATCCAGAAAGGGGAGTATTCAGAGAAATACCTTAACGTAACATTTACTTACTAGTGTCCA	5538
OY		1078	ACACTCTPA 1085	
Db		5537	AGTCTGTGA 5544	
RESULT 11				
AX083744/c				
LOCUS	AX083744	1141 bp	DNA	linear
DEFINITION	Sequence 22 from Patent WO0111061.			PAT 28-FEB-2001
ACCESSION	AX083744			
VERSION	AX083744.1			
KEYWORDS	GI:13185472			
SOURCE	.			
ORGANISM	synthetic construct			
REFERENCE	artificial sequences.			
AUTHORS	Kunst,L. and Clemens,S.			
TITLE	Regulation of embryonic transcription in plants			
JOURNAL	Patent: WO 011061-A 22 15-FEB-2001;			
	UNIVERSITY OF BRITISH COLUMBIA (CA)			

FEATURES					
source		Location/Qualifiers			
		1..1141			
		/organism="synthetic construct"			
		/mol_type="genomic DNA"			
		/_id_xref="taxon:32630"			
promoter		1..1141			
		/note="consensus sequence of A.t., L.a., and B.n. PAEI promoters"			
BASE COUNT	123 a	32 c	42 g	112 t	832 others
ORIGIN					
Query Match		5.8%; Score 63.4; DB 6; Length 1141;			
Best Local Similarity		10.6%; Pred. No. 0.0018;			
Matches		Conservative 247; Misses 239; Indels 3; Gaps 1;			
OY	165 TACAGCAATACGTAAGAAGTTCATCTGTACTGGTAGATGAAGTGAATTGTTGTGCCTC	224			
Dd	::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::	574			
OY	633 TTAMTRMTMGDGGTTVAKKRYKKRNTPTCTTYDDVADMSWWTYANMRCDVTYYTNNTYC	574			
Dd	225 TCACAATACCTTAATGAGCATGTCCAACCGAGAGATCCATCCAGTTCACGGTGAG	284			
OY	::::~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:	573			
Dd	573 KSYASHYWYNMNMWRYSARLNS---SMAFRTIRNNMMMSGVRMRAGTIWMRRHN	517			
OY	285 CGGGAACAATAATTAACATTAACAATTTACGGAAGAAAAGATTAAATAAAGAGACT	344			
Dd	516 NNNDTERTYMWKRRAFTTTYDSMCNAKSMMRGNNRRANKIMWAANNDDAGMDHWY	457			
OY	345 ACAATTAAAGCTATAAACAATTAATGTGCCAAAGTGTAACTGCCATCCGGCCTAAC	404			
Dd	456 MGNMTMMRRAMKMNNMACBRAYCCNNNNRACVHKHKKMYRWTKXMKRLAACNNNBK	397			
OY	405 ACTTAACGACGCTCATTTAAGTGAATATAAAAAAGCGGCTCAGGTGCAAGTTATATT	464			
Dd	396 AMYMFVAAMWSRDITTDMMWVSDEWHNYTDYTMMRAMNNNNNNNMWBCKTTSVMW	337			
OY	465 ATATATTCCTCTGGCGCACTAAAAAATTTGCTTTTGCTGATCTGGATGCTACTCT	524			
Dd	336 MMDHHNHCTIGSNNTWGSABMAAMSMAAGASBNBYIYNMCMRTYKKTWTNNNNNKKA	277			
OY	525 GAAGTTAAGATAAAAAGACATATAGTAGACCTATGAACTTACCATTATAATATCAC	584			
Dd	276 WYFKFYAVMOCNNRRYYDTAVTWBKRKNKYUAWBWYMVGKHNNHMWRRABHRSMN	217			
OY	585 TATTAATAATTAAGTGTAAAGGAATTAATTCACATTAATGTGTACCTCGTCAAAGTGAGCG	644			
Dd	216 MWVCRCRKTYMSHYHAMRYBKMBABYGCCNNMKDRBAHHHCATNNNNMMWYAIIHHM	157			
OY	645 TCGGCGTGCATCTTAAGTTCGCTCCAACGTGTGGGGGCAATATATTTGGAAGAAATTCGT	704			
Dd	156 HKKKAATWNKTBARDDBAHAVKYTWYWRDYCAMCMNAKAKVFPAKMHMYTTRY	97			
OY	705 TGATATG 711				
Dd	96 VSANNTG 90				
RESULT 12					
LOCUS	CEY50E8A	61864 bp	DNA	linear	INV 21-MAY-2003
DEFINITION	Caenorhabditis elegans YAC y50E8a,	complete sequence.			
ACCESSION	AL117200				
VERSION	ALI17200.2	GI:5680305			
KEYWORDS	HTG.				
SOURCE	Caenorhabditis elegans				
ORGANISM	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea; Rhabditiidae; Pelodierinae; Caenorhabditis.				
REFERENCE	1 none.				
AUTHORS	Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium				
JOURNAL	Investigating Biology. The C. elegans Sequencing Consortium Science 282 (5396), 2012-2018 (1998)				
MEDLINE	99069613				


```

PUBMED 9851916
REMARK The C.elegans Sequencing Consortium.
REFERENCE 2 (bases 1 to 61864)
AUTHORS Steward,C.A.
TITLE Direct Submission
JOURNAL Submitted (06-SEP-1999) Nematode Sequencing Project, Sanger
Institute, Hinxton, Cambridge CB10 1SA, England and Department of
Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
jesse@anger.ac.uk or rwnematode.wustl.edu
On Dec 15, 1999 this sequence version replaced g1:5832869.
Coding sequences below are predicted from computer analysis, using
predictions from GeneFinder (P. Green, U. Washington), and other
available information.
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is not the entire insert of clone Y50E8A.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone C4867 is at 61761 in this sequence. The
true right end of clone R08A2 is at 106 in this sequence. The start
of this sequence (1..106) overlaps with the end of sequence Z9853.
The end of this sequence (61761..61864) overlaps with the start of
sequence Z78061.
[991122 d11] : Frameshift detected in sequence. Single nucleotide
removed from the project.
For a graphical representation of this sequence and its analysis
see: http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
name=Y50E8A
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
Location/Qualifiers
1..61864
/organism="Caenorhabditis elegans"
/mol_type="genomic DNA"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="V"
/clone="Y50E8A"
join(1923..2105,2179..2256)
/gene="Y50E8A.1"
join(1923..2105,2179..2256)
/gene="Y50E8A.1"
/standard_name="Y50E8A.1"
/codon_start=1
/product="Hypothetical protein Y50E8A.1"
/protein_id="CAB55044.1"
/db_xref="GI:5832870"
/db_xref="SPRMBL:Q9NAF6"
/translacion="MNFILHILMLSLSLIYTPARPGRNDEKILTGEPLELDGV
DYEDDFVPLNNTKLNENYQYKRLTVSVAGFPKMKRML"
complement(3457..3756)
/gene="Y50E8A.2"
complement(3457..3756)
/gene="Y50E8A.2"
/standard_name="Y50E8A.2"
/codon_start=1
/product="Hypothetical protein Y50E8A.2"
/protein_id="CAB55045.1"
/db_xref="GI:5832871"
/db_xref="GOA:Q9NAF5"
/db_xref="SPRMBL:Q9NAF5"
/translacion="MRLGVDAFCVGLDPKREYVLDVSDAFAFGESTNNDLVEW
TNNPDAARQFREGCGGDMIRKNSQSLTHLFFKSYFFPLVEFFLSNKSFT"
join(4673..4783,5619..5699,5756..5833,5909..6094)
/gene="olig-3"
join(4673..4783,5619..5699,5756..5833,5909..6094)
/gene="olig-3"

FEATURES
source
1..61864
/organism="Caenorhabditis elegans"
/mol_type="genomic DNA"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="V"
/clone="Y50E8A"
join(1923..2105,2179..2256)
/gene="Y50E8A.1"
join(1923..2105,2179..2256)
/gene="Y50E8A.1"
/standard_name="Y50E8A.1"
/codon_start=1
/product="Hypothetical protein Y50E8A.1"
/protein_id="CAB55044.1"
/db_xref="GI:5832870"
/db_xref="SPRMBL:Q9NAF6"
/translacion="MNFILHILMLSLSLIYTPARPGRNDEKILTGEPLELDGV
DYEDDFVPLNNTKLNENYQYKRLTVSVAGFPKMKRML"
complement(3457..3756)
/gene="Y50E8A.2"
complement(3457..3756)
/gene="Y50E8A.2"
/standard_name="Y50E8A.2"
/codon_start=1
/product="Hypothetical protein Y50E8A.2"
/protein_id="CAB55045.1"
/db_xref="GI:5832871"
/db_xref="GOA:Q9NAF5"
/db_xref="SPRMBL:Q9NAF5"
/translacion="MRLGVDAFCVGLDPKREYVLDVSDAFAFGESTNNDLVEW
TNNPDAARQFREGCGGDMIRKNSQSLTHLFFKSYFFPLVEFFLSNKSFT"
join(4673..4783,5619..5699,5756..5833,5909..6094)
/gene="olig-3"
join(4673..4783,5619..5699,5756..5833,5909..6094)
/gene="olig-3"

CDS
gene
/standard_name="Y50E8A.3"
/codon_start=1
/product="C. elegans OIG-3 protein (corresponding sequence
Y50E8A.3)"
/protein_id="CAB55046.1"
/db_xref="GI:5832872"
/db_xref="SPRMBL:Q9NAF4"
/translacion="MSYSPPOPOROYSSITLNDVGVYRSQOQVSPILTKOVNHLAY
EIKNIDPNAMPYAPBEFLKFGDLVHIGGKAVFCVLLGSPRPVCMFENDKLRF
SDVITDPAVDCVLTIPYOMHFRGFTYLCENVEGRATRAQLLELYE"
complement(6492..9248)
/gene="unc-61"
complement(join(6492..6734,6791..7280,7600..7740,
8092..8173,8224..8363,8412..8506,8622..8789,8840..8893,
9069..9248))
/gene="unc-61"
/standard_name="Y50E8A.4a"
/note="cDNA EST YK73392.5 comes from this gene"
/cDNA EST YK73392.3 comes from this gene"
/codon_start=1
/product="C. elegans UNC-61 protein (corresponding
sequence Y50E8A.4a)"
/protein_id="CAB54172.1"
/db_xref="GI:24817603"
/db_xref="SPRMBL:O814C9"
/translacion="MSFETILYASALLFFILVLTFAFLIVRSKSSINLTQTVV
THENPYVTMSOSSKMDIPILPKRFMSIDIEKHLPHPQPPVPHHNOPTHTT
TTSATSSINTTTTSKRTPIAATPASPISKISDHTCKLMONGVDSLEHOLVY
KAVEAFORMLMGVGTGKTTILSLNEMKIDFPCNHELTLERCTKQVAVAGS
IRVKLIVERTAGGDDLDKDSAKVIVDLESQFEYVLOEELKPRMLDYDNVSTHA
CLVIFSPGIGKALDLYLRELAKRVNVIPIYAKSDTCKDLPRKAKTISELSQ
KIDIVTFPPDDETVSTTNEMKNSVPVAVGSDIFVKKNGQVRAQYPMGIVEVEN
ESHCDFVLEALRTNVDERMORTRESLEYENRDRLOMKIGDEGTGPKITKILAO
KHEHODEFSRELTLREEFOKRLDTEDGMKRVESGLAREVEHENTNRESKIDM
EIQTLREERKMLTKYSKILRK"
complement(join(6492..6734,6791..7280,7600..7740,
8092..8173,8224..8363,8412..8506,8622..8795,8840..8860))
/gene="unc-61"
/standard_name="Y50E8A.4b"
/note="contains similarity to Pfam domain: PF00735 (Cell
division protein), Score=415.1, E-value=2.1e-121, N=1"
cDNA EST YK109F1.5 comes from this gene
cDNA EST YK92F8.5 comes from this gene
cDNA EST YK411F2.3 comes from this gene
cDNA EST YK524E2.3 comes from this gene
cDNA EST YK510F2.3 comes from this gene
cDNA EST YK510F2.5 comes from this gene"
/codon_start=1
/product="C. elegans UNC-61 protein (corresponding
sequence Y50E8A.4b)"
/protein_id="CAB60583.1"
/db_xref="GI:6425352"
/db_xref="SPRMBL:Q9U277"
/translacion="MSDIEHKLOLHPHPQPPVPHHNOPTHTNTTSSATSSIN
TTTTSKRTPIAATPASPISKISDHTCKLMONGVDSLEHOLVYKAVEAFORNL
MCYGETGTGKTTILSLNEMKIDFPCNHELTLERCTKQVAVAGSIRVKTGTA
GRGDOLDKDSAKVIVDLESQFEYVLOEELKPRMLDYDNVSTHA
LKALDLYLRELAKRVNVIPIYAKSDTCKDLPRKAKTISELSQKIDIVTFPPD
ETVSTNKNEMKNSVPVAVGSDIFVKKNGQVRAQYPMGIVEVENESHCDFVLE
ALRTVNDERMORTRESLEYENRDRLOMKIGDEGTGPKITKILAOKHEHODEFSR
RELTLREEFOKRLDTEDGMKRVESGLAREVEHENTNRESKIDMELTRELTEEMK
LMTKVSRLRK"
complement(join(9824..9902,9968..10102,10156..10258,
10304..10469,10683..10728,10840..10979,11047..11151))
/gene="Y50E8A.5"
complement(join(9824..9902,9968..10102,10156..10258,
10304..10469,10683..10728,10840..10979,11047..11151))
/gene="Y50E8A.5"
/standard_name="Y50E8A.5"
/codon_start=1
/product="Hypothetical protein Y50E8A.5"
/protein_id="CAB55048.1"

```


CDS

MIKETLSQASIVTDCWGLNEDGCKKIKALMTDPGVKOMRANFESTLPDSAPYEF
 DSIDRMTSPVYIPTQDILHTRVMTRGVHEINFEIGIKERLVDVGQSERKWLSC
 EDDVAVVFCVALISEYDILLHTRVMTRGVHEINFEIGIKERLVDVGQSERKWLSC
 FREKIKHVDLSETPPEYKGRDYERASNYIKERFMQINKTEBOKAIYSHITCATDTNNI
 RVFEAVKDIIETOCVMKAGLYS"
 join(22686, .22743,22820, .22926,23004, .23139,23232, .23558,
 23639, .27180)
 /note="Genid exon scores (in order of location ranges):
 -4.93, 3.57, -1.17, 18.69, 174.64 - GSCJ_ID dd_01857"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AA052224.1"
 /db_xref="GI:28829711"

CDS

complement(join(27340, .29056,29330, .29490))
 /note="Genid exon scores (in order of location ranges):
 218.38, 14.76 - GSCJ_ID dd_01858"
 /codon_start=1
 /product="similar to Rhizobium loti (Mesorhizobium loti).
 Dehydrogenase, succinate semialdehyde dehydrogenase,"

Query Match 4.7%; Score 50.6; DB 3; Length 88549;
 Best Local Similarity 49.8%; Pred. No. 0.56;
 Matches 128; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY	360	TAAACATTTATGTTCAAAAGGTAACTGCCATCCGCCCTAACACTTAACCTCACTCA	419
DB	35734	TACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	35793
QY	420	TTTAACTGTAATAAAAGCGGCTTCAGTCAAGTTAATTAATTAATTAATTAATTAAT	479
DB	35794	TTTCTAATTCAGAGAAATCAACAATATGTAACCTAAATTAATTAATTAATTAATG	35853
QY	480	CGAAGTAAAAATTTGCTTTGTGTGATCTGAGTCTGAGTCTGAGTCTGAGTCTGAGT	539
DB	35854	TGAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	35913
QY	540	AAGAGATATAGTGAAGCTTACCTAATTAATTAATTAATTAATTAATTAATTAATTA	599
DB	35914	ATTACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	35973
QY	600	TAAAGGAATATTTAGA	616
DB	35974	TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	35990

Search completed: August 25, 2003, 18:03:00
 Job time : 398 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 16:45:38 ; Search time 343 Seconds

(without alignments)
8546.909 Million cell updates/sec

Title: US-09-839-894-9

Perfect score: 1086
Sequence: 1 atgaataagattattat.....caagtagtcaaacactctag 1086

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 segs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : N_Geneseq_19jun03:*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1086	100.0	1086	24	AA170763	ETEC CS4 pilus csa
2	1086	100.0	7239	24	AA170780	ETEC CS4 pilus csa
3	298.8	27.5	5798	20	AA84848	CS2 gene cluster.
C 4	49.4	4.5	10279	24	ABL92277	Chemically treated
C 5	49.4	4.5	10279	24	ABL33591	Human immune syste
C 6	49.4	4.5	10279	24	AA22328	Human immune syste
C 7	48.2	4.4	9504	24	ABK28407	Chemically treated
8	47.2	4.3	5139	21	AA170139	DNA transcription
						Plasmodium falcipar

C 9	45.2	4.2	4176	23	ABL10866	Drosophila melanog
C 10	44.8	4.1	13133	24	ABK31231	Signal transductio
C 11	43.8	4.0	6971	24	ABN80061	Human chemically m
C 12	43.4	4.0	2036	23	ABL01847	Drosophila melanog
C 13	43.4	4.0	6080	23	ABL01846	Drosophila melanog
C 14	43.2	4.0	50000	24	ABL55643	AmpeV genome fragm
C 15	42.8	3.9	4985	24	ABG75107	Anopheles gambiæ
C 16	42.8	3.9	6137	24	ABL70127	Chemically treated
C 17	42.6	3.9	6137	24	ABL34450	Human metastasis a
C 18	42.6	3.9	641	24	ABO56694	Human colon cancer
C 19	42.4	3.9	19734	24	ABL33933	Human immune syste
C 20	42.2	3.9	9095	24	ABK28448	DNA transcription
C 21	41.6	3.8	6209	22	AA528751	Genomic sequence #
C 22	41.6	3.8	15548	24	ABL34155	Human immune syste
C 23	41.6	3.8	2365589	24	ABA90521	Chemical sequence o
C 24	41.2	3.8	16228	24	ABL70459	Human gene regulat
C 25	41.2	3.8	16228	24	AA561424	Human gene regulat
C 26	41.1	3.8	13420	24	ABL32917	Human immune syste
C 27	41.1	3.8	16842	22	AA546412	Tumour suppressor
C 28	41.1	3.8	16842	24	ABL70384	Chemically treated
C 29	41.1	3.8	16842	24	AA561336	Human gene regulat
C 30	41.1	3.8	16842	24	ABK31419	Signal transductio
C 31	40.8	3.8	6534	24	ABL32470	Human immune syste
C 32	40.6	3.7	73334	24	ABL92319	Chemically treated
C 33	40.6	3.7	73334	24	ABL34125	Human immune syste
C 34	40.2	3.7	34688	24	ABO67060	Human anglogenesis
C 35	40.2	3.7	6210	24	ABK28471	DNA transcription
C 36	39.8	3.7	9929	22	ABA17906	Human nervous syst
C 37	39.8	3.7	18878	22	AA137441	Human musculoskele
C 38	39.8	3.7	18878	25	ABK60429	cdNA encoding nove
C 39	39.8	3.7	47108	24	ABK31510	Signal transductio
C 40	39.8	3.7	74962	22	AA015256	Human phosphatase
C 41	39.6	3.6	494	23	ABV10021	Human prostate exp
C 42	39.6	3.6	6175	24	AA28373	Human chemically t
C 43	39.6	3.6	7676	24	ABL70409	Chemically treated
C 44	39.6	3.6	7676	24	ABL34598	Human metastasis a
C 45	39.6	3.6	15674	24	ABL70513	Chemically treated

ALIGNMENTS

RESULT 1	AA170763	AA170763 standard; DNA; 1086 BP.
ID	AA170763	
AC	AA170763;	
DT	18-FEB-2002	(first entry)
XX		
DE	ETEC CS4 pilus csaE gene coding region.	
XX		
KW	CS4 pilus; enterotoxigenic; ETEC; csa operon; csaE gene; fimbrial.	
KW	vacine; diarrhoea; antibacterial; antidiarrheic; ds.	
XX		
OS	Escherichia coli.	
XX		
FH	Key	Location/Qualifiers
FT	sig_peptide	1..69
FT	mat_peptide	70..1083
FT		/tag- a
FT		/tag- b
XX		
PN	W0200181582-A2.	
XX		
PD	01-NOV-2001.	
XX		
PF	20-APR-2001; 2001WO-0512914.	
XX		
PR	20-APR-2000; 2000US-198686P.	
XX		
PA	(UYMA-) UNIV MARYLAND BALTIMORE.	
XX		

FT	CDS	/*tag= 1	4196..5281	
FT		/*tag= }		
FT		/note= "csae (claim 31), encodes AAM50343"		
FT	sig_peptide	4196..4264		
FT		/*tag= k		
FT	mat_peptide	4265..5278		
FT		/*tag= l		
FT	CDS	complement (5790..6119)		
FT		/*tag= m		
FT		/note= "(claim 28) csad, encodes AAM50342"		
PN		WO200181582-A2.		
PD		01-NOV-2001.		
XX		20-APR-2001; 2001WO-US12914.		
XX		20-APR-2000; 2000US-198686P.		
XX		(UYMA-) UNIV MARYLAND BALTIMORE.		
XX		Albourn Z, Levine MM, Barry EM;		
DR		WPI: 2002-049280/06		
XX		P-PSDB: AAM50339, AAM50340, AAM50341, AAM50342, AAM50343.		
PT		New nucleotide sequence, useful as immunogenic agent for generating		
PT		immune response against recombinant product of the operon, comprises		
PT		csa operon which encodes enterotoxigenic Escherichia coli-CS4 p111 -		
PS		Disclosure: Page 63-70; 81pp; English.		
XX				
CC		The present sequence is that of the csa operon of enterotoxigenic		
CC		Escherichia coli (ETEC) strain E118181A. The csa operon was		
CC		isolated from a genomic library of this strain. Sequencing revealed		
CC		5 contiguous genes, csaA-csaE, the coding regions of which are		
CC		claimed (see A1170759-63), flanked by 2 insertion elements. The		
CC		csa operon encodes 5 proteins (see AAM50339-43) involved in the		
CC		synthesis of ETEC-CS4 p111. These are the fimbrial structural		
CC		protein (CsaB), the top associated protein (CsaE), a chaperon-like		
CC		protein (CsaA), an usher-like protein (CsaC), and a truncated		
CC		regulatory protein (CsaD). The csa operon has been expressed in		
CC		attenuated Shigella strain CVD1204 guabA, constructing the Shigella		
CC		expressing Csa4 fimbriae vaccine strain CVD1204 (pSA2-CS4). The		
CC		csa operon is useful in the production of recombinant CsaA-CsaE		
CC		polypeptides that are used in claimed immunogenic compositions to		
CC		prevent ETEC colonisation, and hence to protect against diarrhoea.		
XX				
XX		Sequence 7239 BP: 2436 A; 1181 C; 1345 G; 2277 T; 0 other:		
XX				
XX		Query Match	100.0%; Score 1086; DB 24; Length 7239;	
XX		Best Local Similarity	100.0%; Pred. No. 2, 3e-254;	
XX		Matches 1086; Conservative	0; Mismatches	0; Indels
XX			Gaps	0;
QY		1 ATGATAAGATTTTATTTATTTTACATGTTTCTCTCCTAGTACTTTTACATTTTCT	60	
DB		4196 ATGATAAGATTTTATTTATTTTACATGTTTCTCTCCTAGTACTTTTACATTTTCT	4255	
QY		61 GTATCGGAGATTTAAATTTCCCGGAGATGAAAGCATTAATTAATTTTGGCCCGGTAC	120	
DB		4256 GTATCGGAGATTTAAATTTCCCGGAGATGAAAGCATTAATTAATTTTGGCCCGGTAC	4315	
QY		121 AGAAGAGATTTCCCGCCCAACATTAATTTAAATTAACATTAATTAAGCATACGTGAA	180	
DB		4316 AGAAGAGATTTCCCGCCCAACATTAATTTAAATTAACATTAATTAAGCATACGTGAA	4375	
QY		181 AGTCACTCTGTATGATAGATGATGACTTTTATGTTTGTCTCTCACAAATACACTTAAT	240	
DB		4376 AGTCACTCTGTATGATAGATGATGACTTTTATGTTTGTCTCTCACAAATACACTTAAT	4433	
QY		241 GGAGCATCTCAACCGATGAGATCTTAGCAGTTTCATGGTCAAGCGGTGAACAATATTA	300	

Db		4436	GGAGCATGTCCAAACCAAGTGAGAATCTCAGCGATTTCATCGGTCACGGGTGAATAAATA	4495
OY		301	ACATTACAATTTACGGAAAAAGAGTTTAATTA AAAAGAGCTACAAATTTAAAGCCTAT	360
Db		4496	ACATTACAATTTACGGAAAAAGAGTTTAATTA AAAAGAGCTACAAATTTAAAGCCTAT	4555
OY		361	AACAAATTAATGTGCAAAAGTTAACGCCCATCCGGCCCTAACACTTAATCAGCTCAT	420
Db		4556	AAACAAATTAATGTGCAAAAGTTAACGCCCATCCGGCCCTAACACTTAATCAGCTCAT	4615
OY		421	TTTTAACTGTATAATAAAAACGCGGCTTCAGGTGCAAGTTATATTAATATATTCGCTGGC	480
Db		4616	TTTAACTGTATAATAAAAACGCGGCTTCAGGTGCAAGTTATATTAATATATTCGCTGGC	4675
OY		481	GAACTAAAAAAATTTGGCTTTTGGTGGTATCTGGAGTGCCTACTGGAAGTTAAGATAAA	540
Db		4676	GAACTAAAAAAATTTGGCTTTTGGTGGTATCTGGAGTGCCTACTGGAAGTTAAGATAAA	4735
OY		541	AGACGATATAGTAGAGCTATGGAACCTTACCTATAATATCACTATTAAATTAACGTAT	600
Db		4736	AGACGATATAGTAGAGCTATGGAACCTTACCTATAATATCACTATTAAATTAACGTAT	4795
OY		601	AAGGAAATATATTCAGATATGGTTACCTCAGTTCAAAAGTGACGCTCGGCTGCATCTAAC	660
Db		4796	AAGGAAATATATTCAGATATGGTTACCTCAGTTCAAAAGTGACGCTCGGCTGCATCTAAC	4855
OY		661	TTGGCTCCAACTGGTGGGGGACATATATTTGGAAGAAATTCGTGTGATATGTGCTTTTAT	720
Db		4856	TTGGCTCCAACTGGTGGGGGACATATATTTGGAAGAAATTCGTGTGATATGTGCTTTTAT	4915
OY		721	GATGATATATAGTACTTAACAGCAGCTCTTTGAGATAGATTTTCAGGATTAACAATCTTAAA	780
Db		4916	GATGATATATAGTACTTAACAGCAGCTCTTTGAGATAGATTTTCAGGATTAACAATCTTAAA	4975
OY		781	TCTGATGGGAATTTTATCTPAAGAAAAATAATGATGACACCAAGAAATTCGATATCT	840
Db		4976	TCTGATGGGAATTTTATCTPAAGAAAAATAATGATGACACCAAGAAATTCGATATCT	5035
OY		841	TTTGTCACCTCTCTTGGGGGTAAAAATTAACTCCAACAATGGAAGCTCATTTAAATTT	900
Db		5036	TTTGTCACCTCTCTTGGGGGTAAAAATTAACTCCAACAATGGAAGCTCATTTAAATTT	5095
OY		901	GCTGACGACGTTCTCTGGAACAACCTGGAAATGAATTAACGCTGTACCATGCGCAGAA	960
Db		5096	GCTGACGACGTTCTCTGGAACAACCTGGAAATGAATTAACGCTGTACCATGCGCAGAA	5155
OY		961	ATTCAGTGTTCGGGTGTGTGTGGCGTGGACGTTTGCAATTGGATGCAAAAGTGGAAAT	1020
Db		5156	ATTCAGTGTTCGGGTGTGTGTGGCGTGGACGTTTGCAATTGGATGCAAAAGTGGAAAT	5215
OY		1021	CCCCAGGCTGACACATATATGGCTAATTAATTAATTTACTTTCACACCNAAGTAGTCAAACA	1080
Db		5216	CCCCAGGCTGACACATATATGGCTAATTAATTAATTTACTTTCACACCNAAGTAGTCAAACA	5275
OY		1081	CCTTAG 1086 	
Db		5276	CCTTAG 5281	
RESULT 3				
AXX84848				
ID	AXX84848 standard; DNA; 5798 BP.			
XX	AXX84848;			
XX	22-SEP-1999 (first entry)			
DE	CS2 gene cluster.			
KW	CS2 gene cluster; ColA; ColB; ColC; ColD; pilin protein; immunogen;			
KW	enterotoxigenic E. coli; human upper intestine; diarrhoeal disease;			
KW	enteric infection; therapy; ds.			

DR WPI: 2001-656860/75.
XX P-PSDB; ABB66763.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX
PS Claim 1; SEQ ID NO 27080; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB01840-AB16175) and the encoded proteins
CC (AB57737-AB872072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4176 BP; 1186 A; 862 C; 837 G; 1291 T; 0 other;
XX
Query Match 4.2%; Score 45.2; DB 23; Length 4176;
Best Local Similarity 56.6%; Pred. No. 0.45;
Matches 103; Conservative 0; Mismatches 78; Indels 1; Gaps 1;
XX
QY 689 TTGGAAGAAATCTGTGATGATGCTTTATGATGATATAGTACTACAGCAGCTCTT 748
DB 2504 TCGAAGCAAAATTTTAAATTTGTTCTGTATGCGTTACATCAAAAAGACATTTT 2445
QY 749 TGGAGATTAAGATTTCAGATTAACAATCTAAATCTGATGCGAAATTTTATCAAGGAAA 808
DB 2444 AGCAATATCGCATGATGCTAATTAACCGTCTCGTTGAAATATATACATATGAAC 2385
QY 809 TAAATGATGACACCAAGAAATGATATACCT-TTGCACTCTCTGGCGGTAAAGT 867
DB 2384 TAAATTAAGAAATCTAAGACTTTAAATATCTCTAATTAATCTCTGCGCAGAAAT 2325
QY 868 TT 869
DB 2324 TT 2323
XX
RESULT 10
ABK31231/C
ABK31231 standard; DNA; 13133 BP.
XX
ABK31231;
XX
XX 23-APR-2002 (first entry)
XX
XX signal transduction associated gene modified complementary DNA #37.
XX
XX Human: signal transduction associated gene; cytosine methylation state;
XX Cpg island; signal transduction associated disease; solid tumour; cancer;
XX antitumour; cytosine; mutant; ds.
XX
XX Homo sapiens.
XX OS
XX Synthetic.
XX
XX WO200200926-A2.
XX
XX
XX 03-JAN-2002.
XX
XX
XX 29-JUN-2001; 2001WO-EP07472.
XX
XX 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIC-) EPIGENOMICS AG.
XX
XX PA
XX
XX Olek A, Plepenbrock C, Berlin K;

XX
XX WPI: 2002-147896/19.
XX
XX
PT Oligonucleotide for diagnosis and therapy of diseases associated with
PT signal transduction e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with signal transduction -
XX
XX
PS Claim 1; SEQ ID No 74; 24pp; English.
XX
XX
XX The present invention relates to chemically modified DNA sequences of
CC signal transduction associated genes. The DNA sequences are chemically
CC modified using a solution of bisulphite, hydrogen sulphite or
CC disulphite. Also disclosed are oligonucleotides and/or PNA oligomers
CC for detecting the cytosine methylation state (Cpg islands) of these
CC genes, and a method for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with signal transduction.
CC The genomic DNA can be obtained from cells or cellular components which
CC contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,
CC cerebral spinal fluid, tissue embedded in paraffin such as tissue from
CC eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver,
CC histologic object slides, and all their possible combinations. The
CC sequences of the invention are useful for the diagnosis and therapy of
CC diseases associated with signal transduction e.g. solid tumours and
CC cancer. ABK31158-ABK31545 represent chemically pretreated genomic DNA
CC sequences of different genes associated with signal transduction, or
CC their complementary sequences.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 13133 BP; 3676 A; 350 C; 2780 G; 6326 T; 1 other;
XX
Query Match 4.1%; Score 44.8; DB 24; Length 13133;
Best Local Similarity 44.9%; Pred. No. 0.79;
Matches 169; Conservative 0; Mismatches 207; Indels 0; Gaps 0;
XX
QY 4 AATAAGATTTTATTTTATTTTACATGTTTCTCTTCAGTACCTTTTACATTTGCTGTA 63
DB 6047 AAAAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 5988
QY 64 TCGGAGATTAATATTCGCGAGATGAAGCATTAATATTTTGGCGCGGAGACAGG 123
DB 5987 ACATTTAATCCCAAAACTTAATTAATTTCAACAAAATTAATTAATTAATTAATTA 5928
QY 124 AACGAATCTCCGCCCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 183
DB 5927 ATAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5868
QY 184 CATACTCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 243
DB 5867 AATATTTTAAAAAATCACTTAATTTTAAAAATTTTAAAAATTTTAAAAATCACTTA 5808
QY 244 GCATGTCACACAGAGAGATTCCTGACATTCATGCGGAGGAGAAACAATATTAACA 303
DB 5807 ACAATTTTAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5748
QY 304 TTACAATTTACGAAAAAAGAGTTTAATTAATTAATTAATTAATTAATTAATTAATTA 363
DB 5747 ATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 5688
QY 364 CAATTAATGTTCAAAA 379
DB 5687 CTTTACTTAATCAAAA 5672
XX
RESULT 11
ABN80061/C
ABN80061 standard; DNA; 6971 BP.
XX
ABN80061;
XX
XX 15-JUL-2002 (first entry)
XX
XX

QY 786 TGGGAATTATCTAGGAAAT 809
| | | | | | | | |
Db 2369 TAATAATAATAATAATAAT 2346

Search completed: August 25, 2003, 18:08:54
Job time : 347 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 16:47:22 ; Search time 2242 Seconds

(without alignments)
11772.819 Million cell updates/sec

Title: US-09-839-894-9

Perfect score: 1086

Sequence: 1 atgataagatttattat.....caagtagcaaacacttag 1086

Scoring table: IDENTITY_NUC

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Immun DB seq length: 0

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST:*

1: em_estba:*

2: em_esthu:*

3: em_estln:*

4: em_estlu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hnc:*

9: gb_est1:*

10: gb_est2:*

11: gb_hnc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rtd:*

26: em_gss_pmg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	59.4	5.5	1101	29	CNS0100X
2	55	5.1	1056	13	BX415058
3	52	4.8	1200	13	BX415878
4	51.6	4.8	1101	29	CNS00396

Result No.	Score	Query Match	Length	DB ID	Description
C 5	51.2	4.7	1067	14	CD386564
6	51	4.7	1101	29	CNS00FE
7	51	4.7	1225	29	CNS0161D
8	50.8	4.7	1124	13	BX436282
9	50	4.6	1179	13	BX380916
10	49	4.5	1043	29	CNS0145F
11	49	4.5	1061	13	BX437039
12	48.6	4.5	1201	13	BX446296
13	48.4	4.5	1201	13	BX355654
14	48	4.4	1101	29	CNS0039G
15	47.6	4.4	931	14	CD389151
16	47.6	4.4	999	13	BX380865
17	47.6	4.4	1101	29	CNS0145U
18	47.6	4.4	1148	13	BX360752
19	47.4	4.4	925	18	BH152177
20	47.4	4.4	1101	29	CNS000D1
21	47.2	4.3	1101	29	CNS014X1
22	47.2	4.3	1204	29	CNS016E2
23	47	4.3	838	29	CNS00960
24	47	4.3	1101	29	CNS00FMC
25	46.8	4.3	1194	13	BX462185
26	46.8	4.3	536	13	BQ400483
27	46.8	4.3	1101	29	CNS016HF
28	46.6	4.3	408	10	BE224677
29	46.6	4.3	460	12	BE172835
30	46.6	4.3	1201	9	AL532464
31	46.6	4.3	1201	13	BX343343
32	46.4	4.3	812	28	A2529785
33	46.4	4.3	911	28	A2687105
34	46.4	4.3	1092	29	CNS020K7
35	46.4	4.3	1101	29	CNS000B8
36	46.4	4.3	1101	29	CNS00EVL
37	46	4.2	427	13	BX403499
38	46	4.2	1026	13	BX415239
39	45.8	4.2	1201	9	AL536104
40	45.6	4.2	852	29	CNS009AX
41	45.6	4.2	928	29	CNS00DKY
42	45.2	4.2	903	29	CNS017PV
43	45.2	4.2	994	13	BX414650
44	45.2	4.2	1080	29	CNS00EPP
45	45	4.1	1001	29	CNS0155H

ALIGNMENTS

RESULT 1

LOCUS CNS0100X/c 1101 bp DNA linear GSS 26-JUL-1999

DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC

ACCESSION BACN03604 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

VERSION AL098379

KEYWORDS AL098379.1 GI:5609990

SOURCE GSS.

ORGANISM Drosophila melanogaster (fruit fly)

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)

AUTHORS Direct Submission

TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequençage ; BP 191 91006 Evry cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr

JOURNAL

COMMENT

- Web : www.genoscope.cns.fr

Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector

[illegible]

Db	DADWWTWPAADDDWMAADDRMDMAWKMWDMAWMAVGATRADRDWGDGAKRGGARKKRRDRK	799
Oy	CGCGCTTAGGTGGCAAGTTATATTATATATATCTCGCTGGCGGAAGCTAAAAAAATTGGCCT	498
Db	RADDKRDADRDADDAATWTTTWTCTTTTDTDDDKMWKTDTWTWMAADTTWDRDDDDDDADRDAG	739
Oy	TTTGGTGATCTCGGATGCTACTCTGAAGTTAAAGATAAGATAAGATATAGTGAGACC	558
Db	TAGRWRTWKRWMKRRDTRMDDDADDDIARDRRRRRGGDADAGKKTGKRRRRRDRRA	679
Oy	TATGGAAGCTTACACTATAAATATCAGCTATTTAAATTAAGTATGATPAAGGAAATATTCAGATA	618
Db	TWDRDAAWMAAAMWTTTDTDDDKDRRRRGARRRRRTTABAAMDWMTWRAWMAW	619
Oy	TGGTTACTCGTCAAAAGTAGCGCTGGCGGACGCTTAAGCTTACCTGGCTCCAACTGGTGGG	678
Db	DWKRADBDWRMAADTWTIDARADDDMAARWARRRDRARARADRRWTTGKTTATW	559
Oy	GGCAGATATATTGGAAGAATTCCTGTGATATGCTCTTATATGATGAGATATAGTACTAAC	738
Db	TTWABARAAMWMAAATTAATWTTTWT	499
Oy	AGCAGCTTTTGGAGATTAAGATTTCAGGATTAACAATCTTAAATCTGATGGGAAATTTTAT	798
Db	WTAAMAAAAAAMAAWATTTTTTTTTTTTWTWAAWMAWMTATWTTTWTWTWTWTWTWTWTWTWT	439
Oy	CTAAGCAAAATPAAGATGACACCAAGAAATTAAGATATACCTTGCACTTGTC	852
Db	TWTWTWTWMAATTAATTTTWTWTWTWTAAAAAAMWMTWTATATKCCCCCTCC	385

RESULT	5	1067 bp	mRNA	linear	EST 30-MAY-2003
LOCUS	CD386564				
DEFINITION	CD386564	AGNC009T_14286445	NIH_MGC_173	Homo sapiens cDNA 5',	mRNA sequence.
ACCESSION	CD386564				
VERSION	CD386564.1	GI:31222277			
KEYWORDS	EST.				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens (human)				

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 1067)
NIH-MGC <http://mgc.ncl.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Daniela S. Gerhard, Ph.D.

National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgabbs-remail.nih.gov
Tissue Procurement: Dr. Jamie Thompson, University of WI
cDNA Library Preparation: Gina Zastrow-Hayes
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: NDKM74 row: e column: 02
High quality sequence start: 8
High quality sequence stop: 462.
Location/Qualifiers
1.1067

```

/issue_type="Trophoblast"
/issue_host="DH10B Tona"
/clone_id="NH.MGC.173"
/vector="pPONR201; Site_1: attP2; Site_2: attP1;
LIBR.PRIMING - oligo dT; METHOD - full-length enriched
LIBR.PROVIDER - Bradfield"
533 a 85 c 75 g 333 t 41 others
BASE COUNT

```

ORIGIN	Query Match	4.7%;	Score 51.2;	DB 14;	Length 1067;
	Best Local Similarity	44.3%;	Pred. No. 2e+02;		
	Matches 155;	Conservative	0;	Mismatches 195;	Indels 0;
				Gaps 0;	
Oy	290	AAACAATATATACCTTACCATTTCCGGAAAAAGAGAGTTTATATAAAGAGGCTACAA	349		
Db	351	AA	410		
Oy	350	TTAAGGCTATAAACAATTTATGTCAAAAGTGTAACTGCCATCCGGCTAACACTTA	409		
Db	411	AA	470		
Oy	410	ACTGAGCTATTTTACGTATATATAAAACCGCGCTTCAGTGCGAAGTTTATATTATTA	469		
Db	471	ANNN	530		
Oy	470	TTCCGTGCGGCACTAAAAAATTTGCTTTGGTGGATGTCGGAGTCGCTCGAAGT	529		
Db	531	TNTTTTTTTTTTNNAAAAAATTAATTTTTTTTTTATTTTATTAATAAATATATTTTTTTTTT	590		
Oy	530	TAAGAGTAAAAAAGACGATATAGTGAAGCCTATGCACTTACACTATAATATCACTATTA	589		
Db	591	TTATATATATATATATATATTAATANNNTNNAATAATTAATTAANNAATTTATATATATAT	650		
Oy	590	AATTAAGTGAAGGAAATATTCAGATATGCTTACCTCAAGTTCAAAAGT	639		
Db	651	TTAATNTAAAAAATAAT	700		

RESULT 6		
CNS00FE/c		
LOCUS	CNS00FE	1101 bp DNA
DEFINITION	Drosophila melanogaster genome survey sequence TE13 end of BAC: BACR3P18 of Rpci-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.	
ACCESSION	AL071298	
VERSION	AL071298.1	GI:4951138
KEYWORDS	GSS.	
SOURCE	Drosophila melanogaster	(fruit fly)
ORGANISM	Drosophila melanogaster	

REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage

COMMENT	FEATURES	SOURCE
<p>Determination of this BAC-end sequence was carried out as part of collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oseegawa and Aaron Mammossier in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.</p>	<p>Location/Qualifiers</p>	<p>1. .1101</p>

```

FEATURES
  source
    location/Qualifiers
      1..1101
        /organism="Drosophila melanogaster"
        /mol_type="genomic DNA"
        /db_xref="taxon:7227"
        /clone="BACR32P18"
        /clone_id="RPCI-98"

```

with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSP6 vector.

ACCESSION fly), genomic survey sequence.
 VERSION AL103735
 KEYWORDS AL103735.1 GI:5615346
 SOURCE GSS.
 ORGANISM Drosophila melanogaster (fruit fly)
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 1043)
 Genoscope.
 Direct Submission
 Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the European Drosophila Genome Project (EDGP) -
 http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre
 d'Etude du Polymorphisme Humain) with funding provided by a MRC
 project grant. The DNA was prepared from embryos by Alain Bucheton
 and Genevieve Payan. It has been constructed in the vector
 pBelobAC11.
 location/Qualifiers
 1..1043

BASE COUNT	277 a	96 c	121 g	382 t	167 others
ORIGIN					

Query Match	4.5%	Score 49:	DB 29;	length 1043;
Best Local Similarity	37.9%	Pred. No.	4.3e+02	
Matches 136;	Conservative 48;	Mismatches 172;	Indels 3;	Gaps 1

[illegible]

RESULT	11
BX437039	
LOCUS	BX437039
DEFINITION	Homo sapiens THYMS HOME sapiens CDNA clone CS0CAP004TH04
ACCESSION	BX437039
VERSION	BX437039.1
KEYWORDS	EST.
SOURCE	Homo sapiens (human)

ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 1061)
AUTHORS	Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE	Full-length cDNA libraries and normalization
JOURNAL	Unpublished
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr Web : www.genoscope.cns.fr library was constructed by life technologies, a division of Invitrogen. This sequence belongs to sequence cluster 7009.f For more information about this cluster, see http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSOCAP004DD02QP1&cluster=7009.f . Contact : Feng Liang Email : liang@lifetech.com URL : http://fulllength.invitrogen.com/ invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CSOCAP004DD02QP1. Location/Qualifiers 1. 1061
FEATURES	
SOURCE	

BASE COUNT	479 a	77 c	45 g	296 t	164 others
ORIGIN					

Query Match	4.5%	Score 49;	DB 13;	length 1061;
Best Local Similarity	42.2%	Pred. No. 4.3e+02;		
Best local 139; Conservative	17;	Mismatches 173;	Indels 0;	Gaps 0;

[illegible]

RESULT	12
LOCUS	BX446296
DEFINITION	BX446296 Homo sapiens NEUROBLASTOMA cDNA clone
ACCESSION	X1080501ZAD3 3-PRIME, mRNA sequence.
VERSION	BX446296
KEYWORDS	BX446296.1 GI:30784407
SOURCE	EST.
	Homo sapiens (human)

```

ORGANISM      Homo sapiens
REFERENCE     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS       Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE         1 (bases 1 to 1201)
JOURNAL       Ll,W.B., Gruber,C., Jeesee,J. and Polayes,D.
COMMENT       Full-length cDNA libraries and normalization
              Unpublished
              Contact: Genoscope
              Genoscope - Centre National de Sequencage
              BP 191 91006 EVRY cedex - France
              Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
              Library was constructed by Life Technologies, a division of
              Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
              http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
              Faraday Avenue Genoscope sequence ID : XCLDB0012A03P1.

FEATURES
  SOURCE
    1..1201
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="XCLDB0012A03"
      /tissue_type="NEUROBLASTOMA"
      /clone_id="Homo sapiens NEUROBLASTOMA"
      /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
      with a Nott-Oligo(dT) primer. Five prime end enriched,
      double-strand cDNA was digested with Not I and cloned into
      the Not I and EcoRV sites of the pCMVSPORT 6 vector.
      Library was not normalized."

BASE COUNT    663 a      37 c      191 g      116 t      194 others
ORIGIN
Query Match   4.5%; Score 48.6; DB 13; Length 1201;
Best Local Similarity 35.0%; Pred.No. 4.7e+02;
Matches 129; Conservative 53; Mismatches 187; Indels 0; Gaps 0;

QY 11 TTTTATTTATTTTTCATTGTTTTCTCTCAGTACTTTTACATTTGCTGATCGCAG 70
    ||| |||:::||| || |||| | : : : : : : : : : : : : : : : :
DB 152 TTTTATTTTMMWTTTTTTTTTTTTTTTTTTTTTMMWMMWMMWMMWMMWMMWMMW 211
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 71 ATAAATTTCCGGAGATGAAGCATACTAATATTTTGGCCCGGTGACAGGAAGCAT 130
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 212 WWWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMWMMW 271
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 131 CTGCCCAACATPATATTAATACCATTTACGCAATGACATGAAGTCATCTC 190
    | || | | | | | | | | | | | | | | | | | | | | | | | | |
DB 272 ATATATMTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 331
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 191 TGTATGATAGATGACTTTTATGTTGTCTCTCACATATACACTTAATGAGCATGTC 250
    332 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAARRGRBA 391
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 251 CACCCAGGAAATCTCTGACGTTCATCGTCAGCGGTGAACAATATTAACATTACAT 310
    392 AAAAAAAAAAARAAARAAAGGGGGGRRRAARGGGGRRRAAARAAAAAAAAAAAAAA 451
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 311 TTACGGAAAAAAGAGTTTAAATAAAGAGAGCTACAAATTAAGGCTATTAACAATTAT 370
    : |||| : || | || || || | | | | | | | | | | | | | | | |
DB 452 GRBAAAAAAAAAAGGRRAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGAAAAA 511
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 371 TGTTCAAA 379
    : ||||
DB 512 GGGGAAAA 520

RESULT 1.3
LOCUS      BX355654/c
DEFINITION BX355654 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS001002YE12 3-PRIME, mRNA sequence.
ACCESSION  BX355654
VERSION    BX355654.1 GI:30371982
KEYWORDS   EST.
SOURCE     Homo sapiens (human)

```

[illegible]

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 16:50:57 ; Search time 90 Seconds
(without alignments)
5326.023 Million cell updates/sec

Title: US-09-839-894-9

Perfect score: 1086
Sequence: 1 atgaataagatttattat.....caagtagtaaacactctag 1086

Scoring table: IDENTITY_NTC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Minimum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/prodata1/lna/5A_COMB.seq:*
2: /cgn2_6/prodata1/lna/5B_COMB.seq:*
3: /cgn2_6/prodata1/lna/6A_COMB.seq:*
4: /cgn2_6/prodata1/lna/6B_COMB.seq:*
5: /cgn2_6/prodata1/lna/PCTUS_COMB.seq:*
6: /cgn2_6/prodata1/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	298.8	27.5	5798	2	US-08-483-101-1
2	39.8	3.7	74962	4	US-09-685-853A-3
3	39	3.6	10640	4	US-09-417-485D-5
4	38.8	3.6	342	1	US-08-601-198-115
5	38.2	3.5	473	1	US-08-764-100-16
6	38.2	3.5	4970	1	US-08-764-100-14
7	38.2	3.5	4970	1	US-08-764-100-20
8	38.2	3.5	1664976	4	US-08-916-421B-1
9	37.4	3.4	533	4	US-09-495-050A-48
10	36.6	3.4	2671	6	5168051-9
11	36.4	3.4	1020	4	US-09-463-962-1
12	36	3.3	540	4	US-09-404-879A-2
13	36	3.3	540	4	US-09-338-933-2
14	36	3.3	540	4	US-09-215-681-2
15	36	3.3	551	4	US-09-404-879A-137
16	36	3.3	551	4	US-09-338-933-137
17	36	3.3	551	4	US-09-215-681-137
18	36	3.3	15016	4	US-09-601-198-60
19	36	3.3	319608	4	US-09-539-33D-1
20	36	3.3	319608	4	US-09-679-409-1
21	35.6	3.3	1689	1	US-07-991-867B-41
22	35.6	3.3	1689	2	US-08-544-332-41
23	35.6	3.3	1689	4	US-09-370-861A-41
24	35.6	3.3	1947	4	US-09-370-861A-74
25	35.6	3.3	8457	1	US-07-991-867B-1
26	35.6	3.3	8457	2	US-08-544-332-1
27	35.6	3.3	8457	4	US-09-370-861A-1

28	35.6	3.3	19124	2	US-08-487-826B-13	Sequence 13, Appl
29	35.4	3.3	2826	4	US-08-624-655A-1	Sequence 1, Appl1
30	35.4	3.3	162450	4	US-09-345-882-1	Sequence 1, Appl1
31	35.4	3.3	169998	4	US-09-676-610B-24	Sequence 24, Appl
32	35.4	3.3	197496	4	US-09-877-177A-10	Sequence 10, Appl
33	35	3.2	1491	4	US-09-134-001C-866	Sequence 866, App
34	35	3.2	11770	4	US-08-961-527-172	Sequence 172, App
35	34.8	3.2	337	2	US-09-032-684-11	Sequence 11, Appl
36	34.8	3.2	612	4	US-09-107-532A-1950	Sequence 1950, Ap
37	34.6	3.2	4011	1	US-08-121-057-3	Sequence 3, Appl1
38	34.6	3.2	4011	2	US-08-509-187D-3	Sequence 3, Appl1
39	34.6	3.2	4011	2	US-09-121-396-3	Sequence 3, Appl1
40	34.6	3.2	4011	5	PCT-US93-09704A-3	Sequence 3, Appl1
41	34.6	3.2	4323	4	US-09-620-312D-134	Sequence 134, App
42	34.6	3.2	4838	3	US-08-852-629-15	Sequence 15, Appl
43	34.6	3.2	1664976	4	US-08-916-421B-1	Sequence 1, Appl1
44	34.4	3.2	195	4	US-09-328-352-1980	Sequence 1980, Ap
45	34.4	3.2	801	4	US-09-134-001C-1507	Sequence 1507, Ap

ALIGNMENTS

RESULT 1
US-08-483-101-1
Sequence 1, Application US/08483101
Patent No. 5932715
GENERAL INFORMATION:
APPLICANT: Scott, June R.
APPLICANT: Froehlich, Barbara
TITLE OF INVENTION: CS2 Proteins and Coding Sequences
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee and Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,101
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Farber, Donna M.
REGISTRATION NUMBER: 33878
REFERENCE/DOCKET NUMBER: 6-95
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5798 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: 499..1215
FEATURE:
NAME/KEY: sig-peptide
LOCATION: 499..552

```

FEATURE:
NAME/KEY: mat_peptide
LOCATION: 553..1212
FEATURE:
NAME/KEY: CDS
LOCATION: 1255..1767
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 1255..1323
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 1324..1764
FEATURE:
NAME/KEY: CDS
LOCATION: 1836..4436
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 1836..1913
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 1914..4433
FEATURE:
NAME/KEY: CDS
LOCATION: 4451..5545
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 4451..4504
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 4505..5542
US-08-483-101-1

```

```

Query Match 27.5%; Score 298.8; DB 2; Length 5796;
Best Local Similarity 58.5%; Pred. No. 7,7e-70;
Matches 566; Conservative 0; Mismatches 387; Indels 15; Gaps 2;

```

```

QY 127 GAATCTGCCCAACATATATTAATACATATATACAGCTACAGTAAAGTCT 186
DB 4583 GCATGCTATCTGCTCATATATATTTCTGACACATGTTGGTTACATATAAGTAC 4642
QY 187 ACTCTGTATGATAGATGACCTTTTATGTTGCTCTTCACAAATACACTTAATGAGCA 246
DB 4643 TCTCTTTTGACAGATGACGTTTTATGATATGATCATCAAGATGATCTAAAGTGA 4702
QY 247 TGTCCACCATGAGAAATCTAGCATTCATCGCTGACGGGTGAACAAATATTAACATTA 306
DB 4703 TGTCCGACGAGGAAAGTCCCA-----AATCCTCTCAAGGGGAGACTAATATTAAGCTA 4756
QY 307 CAATTTAGGAAAAAGAAAGTTTAATAAAGAGAGCTACAAATTAAGGGCTATTAACAA 366
DB 4757 ATATTTACTGAAAGAAAGAGTGTGGCCAGAAAAACATTTAACTTAAAGGATATTAAGCA 4816
QY 367 TTATTTGTCAAAAGTGTAACTGC-----CCATCGCGGCTTAACACTTAACACAGT 417
DB 4817 TTTTATATGATATGATATGATGATGATCATTTATGCGATTAATAATGAATCTCATCTCAT 4876
QY 418 CATTTTAAGTGAATAAAAACCGGCTTCAGCTGCAAGTTTAAATTAATATATTCCTGCT 477
DB 4877 ACTGTTAATGTGTAGTTCAATTCACAAAGAGAGTATGATTTCACTTATATATCCACAA 4936
QY 478 GGGCAATAAAAAATTTGGCTTTTGGTATCTGGATGCTAGCTGAGATAGAGTA 537
DB 4937 GGTGAAATGATGGCTTCTTAATCTGAGGATATATGGAGGCAACACTAGAGTTAGAGTC 4996
QY 538 AAAAGACATATAGTACAGCTATGAACTTACACTATATATATCACTATTAATTAAT 597
DB 4997 AAAAGCATATGACTATATATATGATGATGATGATGATGATGATGATGATGATGAT 5056
QY 598 GATTAAGGAAATATTCAGATATGATGATGATGATGATGATGATGATGATGATGATGAT 657
DB 5057 GACAAAGGAAATATTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5116
QY 658 AACTGGCTCAACAGTGTGGGGGACATATATGGAAGAAATTCGTGATATGTCCTT 717

```

```

DB 5117 AATTACGTCCTGAAGATATAGTAAATATCTGCTAGTAAAGTCTGATGATGCTC 5176
QY 718 TATGATGATATAGTACTAAGACGACGCTTTGGATATAGATTCAGATTAACAATCCT 777
DB 5177 TATGATGCTATAGTACATATGATGATGATGATGATGATGATGATGATGATGATGAT 5236
QY 778 AATCTGATGGAAATTTATCTAAGAAATTAATGATGACACCAAGAAATGATGAT 837
DB 5237 ACAGAAATATATGATATATATCTTATTAATACTGAGACCATTAATAAATGTCATAT 5296
QY 838 ACTTGTCACTTCTTGGCGGTTAAAGTTTAACTCCACAAATGAGATCAATTAAT 897
DB 5297 AACTTCTCTTCTTTTGGAGAGACAGATTTTATCCAAATTAAGGAGAGCTTTTACT 5356
QY 898 ATTGCTGACGACGCTTCTGTAACAAACATGATATGATTAAGCTGATGATGATGAT 957
DB 5357 ATTAATGATATCTGCTCATTTGTTTAAACTGGAATGCTATTAAGTCTGATATCTTACCA 5416
QY 958 GAAATCAGTGTCCGCTGTGTGTGCTGACGCTTTCATTAATGATGATGATGATGAT 1017
DB 5417 CAGATTAATGATATTCAGATATGATGATGATGATGATGATGATGATGATGATGATGAT 5476
QY 1018 AATCCGAGCTGGACAAATATATGATGATATTAATGATGATGATGATGATGATGAT 1077
DB 5477 AATCCGAGAGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5536
QY 1078 AACTCTTA 1085
DB 5537 AGTCTGTA 5544

```

```

RESULT 2
US-09-685-853A-3/C
; Sequence 3, Application us/09685853A
; Patent No. 6479270
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming et al.
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
; FILE REFERENCE: CLO00871
; CURRENT APPLICATION NUMBER: US/09/685,853A
; CURRENT FILING DATE: 2002-05-06
; PRIOR APPLICATION NUMBER: 60/182,194
; PRIOR FILING DATE: 2000-02-14
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO. 3
; LENGTH: 74962
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(74962)
; OTHER INFORMATION: n = A,T,C or G
US-09-685-853A-3

```

```

Query Match 3.7%; Score 39.8; DB 4; Length 74962;
Best Local Similarity 46.8%; Pred. No. 1.1;
Matches 125; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

```

1 APPLICATION NUMBER: US/08/214,066
 2
 3 FILING DATE:
 4 APPLICATION NUMBER: US 08/032,235
 5
 6 FILING DATE: 17-MAR-1993
 7
 8 APPLICATION NUMBER: GB 9206016.9
 9
 10 FILING DATE: 19-MAR-1992
 11
 12 ATTORNEY/AGENT INFORMATION:
 13
 14 NAME: No. 57737001s, Allen E.

REGISTRATION NUMBER: 34,490
 REFERENCE/DOCKET NUMBER: 137-1061
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 354-3592
 TELEFAX: (415) 857-1125
 INFORMATION FOR SEQ ID NO: 16:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 473 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-764-100-16

Query Match 3.5%; Score 38.2; DB 1; Length 473;
 Best Local Similarity 43.8%; Pred. No. 0.5;
 Matches 166; Conservative 0; Mismatches 213; Indels 0; Gaps 0;

```

420 TTTTACGTATATAAAGCGGCTTCAGTGCAGTTATATATATATATCCGCTGG 479
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
47 TTGAATGTTTAAAGTAAATTAACCAAGATATAATATATATATATATATAGA 106
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
480 CGACTTAAATAATTTGCTTTTGTGTGATCTGGATGCTACTGTAAGTTAAG 539
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
107 AGTATATAATATATATATATATATATATATATATATATATATATATAGA 166
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
540 AAGACGATATAGTGAACCTATGAACTTACATATAATATCAGTATTAATTA 599
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
167 AAAAAATTAACAAAAACAAAAACAAAAACAAAAACAAAAACAAAAATGAAA 226
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
600 TAAAGGAATATTCAGATATGTTACCTGACGTTCAAAAGTGAGCGTGGGAT 659
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
227 AAGTTGAAAAAACCAAAAAATTTTGTGTAATTAATTAAGGTCGGCCAGATT 286
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
660 CTTCGCTCCAACTGTGGGGCACAATATTTGAAGAAATCTGTGATATGCTTT 719
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
287 TCTAAGACCTTTTATTTGTTTATACATTTTATTTGTTGATTTTATTTT 346
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
720 TGATGATATATAGTACTAACAGAGCTCTTGGAGTAAAGTTTCAAGATAC 779
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
347 TTATTTTATATTTTATATATAGTTGCTTATTAACCTTATTTAGAACAA 406
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
780 ATCTGATGGAATTTTAT 798
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
407 ATTGATTACATCATCTCT 425
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||

```

SEQUENCE 6
 08-764-100-14
 Sequence 14, Application US/08764100
 Patent No. 5773700
 GENERAL INFORMATION:
 APPLICANT: van Grinsven J., Martinus O.
 APPLICANT: De Haan, Petrus T.
 APPLICANT: Gielen L., Johannes J.
 APPLICANT: Peters, Dirk
 APPLICANT: Goldbach, Robert W.
 TITLE OF INVENTION: Improvements in or Relating to Organic
 NUMBER OF SEQUENCES: 27
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sandoz Agro, Inc
 STREET: 975 California Avenue
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/764,100

FILING DATE: 06-DEC-1996
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/214,064
 FILING DATE:
 APPLICATION NUMBER: US 08/032,235
 FILING DATE: 17-MAR-1993
 APPLICATION NUMBER: GB 9206016.9
 FILING DATE: 19-MAR-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: No. 5773700-15; Allen E.
 REGISTRATION NUMBER: 34,490
 REFERENCE/DOCKET NUMBER: 137-1061
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 354-3592
 TELEFAX: (415) 857-1125
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4970 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-764-100-14

Query Match 3.5%; Score 38.2; DB 1; Length 4970;
 Best Local Similarity 43.8%; Pred. No. 1.1;
 Matches 166; Conservative 0; Mismatches 213; Indels 0; Gaps 0;

```

420 TTTTACGTATATAAAGCGGCTTCAGTGCAGTTATATATATATATCCGCTGG 479
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
1044 TTGAATGTTTAAAGTAAATTAACCAAGATATAATATATATATATATAGA 1103
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
480 CGACTTAAATAATTTGCTTTTGTGTGATCTGGATGCTACTGTAAGTTAAG 539
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
1104 AGTATATAATATATATATATATATATATATATATATATATATATAGA 1163
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
540 AAGACGATATAGTGAACCTATGAACTTACATATAATATCAGTATTAATTA 599
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
1164 AAAAAATTAACAAAAACAAAAACAAAAACAAAAACAAAAACAAAAATGAAA 1223
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
600 TAAAGGAATATTCAGATATGTTACCTGACGTTCAAAAGTGAGCGTGGGAT 659
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
1224 AAGTTGAAAAAACCAAAAAATTTTGTGTAATTAATTAAGGTCGGCCAGATT 1283
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
660 CTTCGCTCCAACTGTGGGGCACAATATTTGAAGAAATCTGTGATATGCTTT 719
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
1284 TCTAAGACCTTTTATTTGTTTATACATTTTATTTGTTTGTGATTTTATTT 1343
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
720 TGATGATATATAGTACTAACAGAGCTCTTGGAGTAAAGTTTCAAGATAC 779
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
1344 TTATTTTATATTTTATATATAGTTGCTTATTAACCTTATTTAGAACAA 1403
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
780 ATCTGATGGAATTTTAT 798
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
1404 ATTGATTACATCATCTCT 1422
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||

```

SEQUENCE 7
 US-08-764-100-20/c
 Sequence 20, Application US/08764100
 Patent No. 5773700
 GENERAL INFORMATION:
 APPLICANT: van Grinsven J., Martinus O.
 APPLICANT: De Haan, Petrus T.
 APPLICANT: Gielen L., Johannes J.
 APPLICANT: Peters, Dirk
 APPLICANT: Goldbach, Robert W.
 TITLE OF INVENTION: Improvements in or Relating to Organic
 NUMBER OF SEQUENCES: 27
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sandoz Agro, Inc
 STREET: 975 California Avenue

CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,100
FILING DATE: 06-DEC-1996
CLASSIFICATION: 800
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/08/214,064
FILING DATE:
APPLICATION NUMBER: US 08/032,235
FILING DATE: 17-MAR-1993
APPLICATION NUMBER: GB 9206016.9
FILING DATE: 19-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 57737001s, Allen E.
REGISTRATION NUMBER: 34,490
REFERENCE/DOCKET NUMBER: 137-1061
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 354-3592
TELEFAX: (415) 857-1125
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 4970 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-764-100-20

Query Match 3.58; Score 38.2; DB 1; Length 4970;
Best Local Similarity 43.88; Pred. No. 1.1;
Matches 166; Conservative 0; Mismatches 213; Indels 0; Gaps 0;

QY 420 TTTTAACTGTAATATAAAGCGGCTTCAGGTGCAAGTTTATTTATATATATCTGCTGG 479
DB 3927 TTGAATGTTTAACTATAAATAAAGCAAGTATAAATACTATATATATATATATATAGA 3668
QY 480 CGAAGTAAAAAATTTGGCTTTGGTGTATCTGGATGCTACTCTGGAAGTTAAGAGTAAA 539
DB 3867 AGATATAAATAATATATATATGTTTGTGTTTAAAAACAATCAAAAACCAAAAAGAAAAAG 3808
QY 540 AAGACGATATAGTGAACCTATGGAATCTACATATATATATATATATATATATATCTGA 599
DB 3807 AAAAAATTAACAAAAAACAACAAAAACAAAAACAAAAACAAAAAAGCAAAAAATAGAAAA 3748
QY 600 TAAGGGAATATATCAGATATGTTTACCTCAGTTCAAAAGTGAAGCGGCTGCATCTTAA 659
DB 3747 AAGTTGAAAAAACAACAAAAAATTTTTTTTGTAAATATAAAGCTCCGCCAGATTTGG 3688
QY 660 CTGCGCTCAACTGGTGGGGGACATATATATGGAAGAAATCTGTTGATATGCTTTTA 719
DB 3687 TCTAAGACCTTTTATTTGTTTATTTTATACATTTTATTTGTTTGTGATTTTATTTTA 3628
QY 720 TGATGATATAGTACTACAGCAGCTCTTGGAGATAGTTTCAGGATACCAATCTTAA 779
DB 3627 TTTATTTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 3568
QY 780 ATTCGATGGAATTTTAT 798
DB 3567 ATTGATTAACAATCATTTCT 3549

RESULT 8
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:

APPLICANT: Bult et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methano
Patent No. 6503729
TITLE OF INVENTION: jannaschii
FILE REFERENCE: PB275
CURRENT APPLICATION NUMBER: US/08/916,421B
CURRENT FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc_feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g

```

NAME/KEY: misc_feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (741584)..(741584)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature

```

```

LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1348473)..(1348473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1664854)..(1664854)
OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

```

```

Query Match 3.5%; Score 38.2; DB 4; Length 1664976;
Best Local Similarity 51.5%; Pred. No. 8.2;
Matches 88; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

```

```

QY 747 TTTCGAGATTAAGATTTCAGATTAACAATCTTAATCTGATGGAATTTTATGAGAA 806
DB 1486909 TGTGAATTTGATTAAGAAACAAATCCGTTTATGTTAAATTCCTATGAAAACT 1486968
QY 807 AATAATGATGACCAAGAAATTCATTAATCTTGTCTTCTTGGCGGTAAG 866
DB 1486969 CTTCAGCTATTAACCAAGATGATGATCTATTTGAATTAAGATTTGGAGCGGCTT 1487028
QY 867 TTTCACCTCAACCAATGGAAGTCATTAATATTCGTACCGACGCTTCT 917
DB 1487029 TTTAATAAACAGATATCTCTTTTATGAAAAATACCAAAATCATCT 1487079

```

```

RESULT 9
US-09-495-050A-48
Sequence 48, Application US/09495050A
Patent No. 6492505
GENERAL INFORMATION:
APPLICANT: Roopa, Reddy
APPLICANT: Guegler, Karl, J.
APPLICANT: Au-Young, Janice
TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING MEMBRANE-ASSOCIATE
FILE REFERENCE: PA-0013 US
CURRENT APPLICATION NUMBER: US/09/495.050A
PRIOR FILING DATE: 2000-01-31
PRIOR APPLICATION NUMBER: 60/718,318
NUMBER OF SEQ ID NOS: 305
SOFTWARE: PERL Program
SEQ ID NO 48
LENGTH: 533
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. 6492505 949387CT1
US-09-495-050A-48

```

```

Query Match 3.4%; Score 37.4; DB 4; Length 533;
Best Local Similarity 48.4%; Pred. No. 0.84;
Matches 104; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

```

LOCUTION: (1)..(1020)
US-09-463-962-1
Query Match 3.4%; Score 36.4; DB 4; Length 1020;
Best Local Similarity 52.7%; Pred. No. 1.9;
Matches 79; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
QY 140 AACATATATATTAATACCATTTACACATACAGTGAAGTCATCTCTGTATGATA 199
D 132 AAGTATATGATTTCTTGTGAAAAATGCTAAGAGAGATTTAAATATCCTCACACAAAAAA 191
QY 200 GGATGACTTTTATGTTTCTTCTTCACAAATACACTTAATGGAGCATGCCAACCGTG 259
D 192 TGATTAATATATGATTAATATACATATGTCATACATGCTCAATTTAGCCATTCATATGTA 251
QY 260 AGATCTCAGCAGTTCATCGTCAGCGTGGAACCAATATATACATTTACATTTACGAAA 319
D 252 TGCATGTTTCAAAAACATGTCGTACATATATATATATATTTTGTGAGTGGAGA 311
QY 320 AAGAGATTAT 354
D 312 TAAATAGAGTGTAT 346
RESULT 10
5168051-9/c
PATENT NO. 5168051
APPLICANT: DERYNCK, RIK M.A.; GOEDDEL, DAVID V.
TITLE OF INVENTION: NUCLEIC ACID ENCODING TGF-B ITS USES
NUMBER OF SEQUENCES: 21
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/389,929
FILING DATE: 04-AUG-1989
SEQ ID NO: 9
LENGTH: 2671
Query Match 3.4%; Score 36.6; DB 6; Length 2671;
Best Local Similarity 51.5%; Pred. No. 2.4;
Matches 84; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
QY 11 TTTATTTATTTTACATTTTCTCTCAGACTTTTACATTTGCTGATCCGAG 70
D 2638 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 2579
QY 71 ATTAATTTCCCGAGATGAAGCATATATATTTTGGCCCGGAGAGAAAGCAT 130
D 2578 ACCTTAATATTAAGCCCTTTTGTGTTAGAAACATCTCCAGAAAGATGATCCAG 2519
QY 131 CTTCCTCCCAACAT 173
D 2518 ATTCCTCCCAACAT 2476
RESULT 11
US-09-463-962-1
Sequence 1, Application US/09463962
Patent No. 6413754
GENERAL INFORMATION:
APPLICANT: FAYE, GERARD
APPLICANT: VALAY, JEAN-GABRIEL
APPLICANT: MANN, CARL
APPLICANT: THORET, JEAN-YVES
TITLE OF INVENTION: KINASE ACTIVATING DEPENDENT CYCLIN PROTEIN KINASES AND THEIR USES
FILE REFERENCE: 0846-0536-0XCT
CURRENT APPLICATION NUMBER: US/09/463,962
CURRENT FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: PCT/FR98/01788
PRIOR FILING DATE: 1998-08-11
PRIOR APPLICATION NUMBER: FR97/10287
PRIOR FILING DATE: 1997-08-12
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 1020
TYPE: DNA
ORGANISM: Candida albicans
FEATURE:
NAME/KEY: CDS

LOCUTION: (1)..(1020)
US-09-463-962-1
Query Match 3.4%; Score 36.4; DB 4; Length 1020;
Best Local Similarity 52.7%; Pred. No. 1.9;
Matches 79; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
QY 679 GGCATATATATGGAAGAAATTCGTGATATGCTTTATGATGATATATGATATGATATGATATGAT 738
D 790 GGTACACCACTTTTAACTGATTTTGAAGATGATATATTTTGTGATGATATATATATATGAA 849
QY 739 AGCAGCTTTTGGAGATTAAGATTTTACAGATTAACATCTTAAATCTGATGGAATTTTAT 798
D 850 AACTGCAATTTTAAAAATTCATTTACAAAATATCTTAAGAAAGATGGATATATAT 909
QY 799 CTAAAGAAATTAATGATGACCAAGAA 828
D 910 TTACCTGATGCAATGATGATTTGATGAA 939
RESULT 12
US-09-404-879A-2
Sequence 2, Application US/09404879A
Patent No. 648546
GENERAL INFORMATION:
APPLICANT: Mitcham, Jennifer L.
APPLICANT: King, Gordon E.
APPLICANT: Algate, Paul A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.462C2
CURRENT APPLICATION NUMBER: US/09/404,879A
CURRENT FILING DATE: 1999-09-24
NUMBER OF SEQ ID NOS: 393
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 540
TYPE: DNA
ORGANISM: Homo sapien
US-09-404-879A-2
Query Match 3.3%; Score 36; DB 4; Length 540;
Best Local Similarity 46.6%; Pred. No. 2;
Matches 151; Conservative 0; Mismatches 170; Indels 3; Gaps 1;
QY 318 AAAAGAGTTTAT 377
D 66 AAGATGCAATTTAAATATGAGTTTCTTCACTTTATCTGAGACAAAGTATCCATTA 125
QY 378 AAGTTTAACTGCGCCATCGGCTTAACCTTAACCTTATCTGATATATATATATATATATATATATAT 437
D 126 TTTATGTCAGAAAGATGATTAACCTGCTTAAGAGCTTACGAAGCTATGGAGAG 185
QY 438 CGGCGCTTCAGGTCGAAGTTTAT 497
D 186 GTTGGCAGCAAGCAATTTAT 245
QY 498 TTTTGG--TGATATCGGGATGCTACTGGAAGTAAAGTAAAGAAAGAGATATATGTA 554
D 246 TTTCTGATGGAAGCTATATGAGCTTATTTGAAATGAGAGTTTACGAAGGCAATGAGACC 305
QY 555 GACCTATGAACTTACAT 614
D 306 GCGAGACTGTGTATGCAATTAATGAGAGCTTTATATGAACTATATATATATATATATATATATATAT 365
QY 615 GATATGTTTACCTCAATGTTCAAG 638
D 366 AGCAGGTTTACATGATGAAAAAG 389
RESULT 13
US-09-338-933-2
Sequence 2, Application US/09338933

```

; Patent No. 6488931
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer Lynn
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF
; TITLE OF INVENTION: OVARIAN CANCER
; FILE REFERENCE: 210121.46201
; CURRENT APPLICATION NUMBER: US/09/338, 933
; CURRENT FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 312
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 540
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-338-933-2

```

Query Match 3.3%; Score 36; DB 4; Length 540;

Best Local Similarity 46.6%; Pred. No. 2;

Matches 151; Conservative 0; Mismatches 170; Indels 3; Gaps 1;

```

QY 318 AAAAAGAGTTTAAATTAAGAGAGCTTAAAGCTTAAAGCTTAAATTTGTTCA 377
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 66 AAGATGCAATTAAATATGGTTATTTCACTTTTATCTGAGGACAAATCCATTAA 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 378 AAGTTTAAGTCCCATCCGCTTAACACTTACTGCTCATTTTAACTGTAATAAAA 437
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 126 TTATTTGTGCAGAAAGATGTAATACCTGCTTAAGAAGCTTACAGAACTATGGGAG 185
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 438 CGCGGCTTCAGTGCAGTTTATTTATTTCTGCTGCGCAACTTAAATTTGCC 497
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 186 GTTGGACGACAAACAATTTGAACATTTAAATCACTTATGACAGTAAATGAGCC 245
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 498 TTTTGG---TGGTATCTGGGATCTACTCTGAAGTAAAGTAAAGACATATAGTGA 554
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 246 TTTCGACAGGAACTTATGAGCTTATGAAAGGACAGTTTGCAAGGACATGAGACC 305
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 555 GACCTATGAACTTACATATATATCTATTAATTAAGTAAAGGAAATTTTCA 614
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 306 GGCAGACTGTCTATGCGCAATTAAGAAGCTTTAAATGAACTTATATGATGTGTAA 365
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 615 GATATGTTACCTCAGTTCAAAAG 638
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 366 AGCAGGTTACATGATGAAAAAG 389
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

ULN 14
09-215-681-2
; Sequence 2, Application US/09215681A
; Patent No. 6528253
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Frudakis, Tony N.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
; TITLE OF INVENTION: OF OVARIAN CANCER
; FILE REFERENCE: 210121.463
; CURRENT APPLICATION NUMBER: US/09/215, 681A
; CURRENT FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 310
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 540
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-215-681-2

```

Query Match 3.3%; Score 36; DB 4; Length 540;

Best Local Similarity 46.6%; Pred. No. 2;

Matches 151; Conservative 0; Mismatches 170; Indels 3; Gaps 1;

```

QY 318 AAAAAGAGTTTAAATTAAGAGAGCTTAAATTAAGGCTTAAACATTTTGTCAA 377
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 66 AAGATGCAATTAAATATGGGTTATTTTCACCTTTTATCTGAGGACAAATCCATTAA 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 378 AAGTTTAAGTCCCATCCGCTTAACACTTACTGCTCATTTTAACTGTAATAAAA 437
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 126 TTATTTGTGCAGAAAGATGTAATACCTGCTTAAAGCTTACAAAGCTATGGAGAG 185
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 438 CGCGGCTTCAGTGCAGTTTATTTATTTATTTCTGCTGCGCAACTTAAATTTGCC 497
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 186 GTTGGACGACAAACAATTTGAACATTTAAATCACTTATGATGACAGTAAATGAGCC 245
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 498 TTTTGG---TGGTATCTGGGATCTACTCTGAAGTAAAGTAAAGACATATAGTGA 554
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 246 TTTCGACAGGAACTTATGAGCTTATGAAAGGACAGTTTGCAAGGACATGAGACC 305
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 555 GACCTATGAACTTACATATATATCTATTAATTAAGTAAAGGAAATATTTCA 614
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 306 GGCAGACTGTCTATGCGCAATTAAGAAGCTTTAAATGAACTTATATGATGTGTAA 365
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 615 GATATGTTACCTCAGTTCAAAAG 638
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 366 AGCAGGTTACATGATGAAAAAG 389
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 15

US-09-404-879A-137

; Sequence 137, Application US/09404879A

; Patent No. 6468546

; GENERAL INFORMATION:

; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Algate, Paul A.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER

; FILE REFERENCE: 210121.46202

; CURRENT APPLICATION NUMBER: US/09/404, 879A

; CURRENT FILING DATE: 1999-09-24

; NUMBER OF SEQ ID NOS: 393

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 137

; LENGTH: 551

; TYPE: DNA

; ORGANISM: Homo sapien

US-09-404-879A-137

Query Match 3.3%; Score 36; DB 4; Length 551;

Best Local Similarity 46.6%; Pred. No. 2;

Matches 151; Conservative 0; Mismatches 170; Indels 3; Gaps 1;

```

QY 318 AAAAAGAGTTTAAATTAAGAGAGCTTAAATTAAGGCTTAAACATTTTGTCAA 377
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 63 AAGATGCAATTAAATATGGGTTATTTTCACCTTTTATCTGAGGACAAATCCATTAA 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 378 AAGTTTAAGTCCCATCCGCTTAACACTTACTGCTCATTTTAACTGTAATAAAA 437
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 123 TTATTTGTGCAGAAAGATGTAATACCTGCTTAAAGCTTACAAAGCTATGGAGAG 182
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 438 CGCGGCTTCAGTGCAGTTTATTTATTTATTTCTGCTGCGCAACTTAAATTTGCC 497
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 183 GTTGGACGACAAACAATTTGAACATTTAAATCACTTATGATGACAGTAAATGAGCC 242
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 498 TTTTGG---TGGTATCTGGGATCTACTCTGAAGTAAAGTAAAGACATATAGTGA 554
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 243 TTTCGACATGGAACCTTATGAGCTTATGAAAGGACAGTTTGAAGGACATGAGACC 302
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 555 GACCTATGAACTTACATATATATCTATTAATTAAGTAAAGGAAATATTTCA 614
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 303 GGCAGACTGTCTATGCGCAATTAATGAAGCTTTAAATGAACTTATATGATGTGTAA 362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 615 GATATGTTACCTCAGTTCAAAAG 638
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 363 AGCAGGTTACATGATGAAAAAG 386
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: August 25, 2003, 18:48:16
Job time : 97 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 25, 2003, 18:03:07 ; Search time 289 Seconds
(Without alignments)
8448.117 Million cell updates/sec

Title: US-09-839-894-9

Perfect score: 1086
Sequence: 1 atgaataagatttattat.....caagtagtcaacactctag 1086

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1517243 seqs, 1124081882 residues

Total number of hits satisfying chosen parameters: 3034486

Imm DB seq length: 0
Imm DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications, NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1086	100.0	1086	US-09-839-894-9	Sequence 9, Appl
2	1086	100.0	7239	US-09-839-894-27	Sequence 27, Appl
3	49.4	4.5	10279	US-10-311-455-1564	Sequence 1564, Ap
4	48.2	4.4	9504	US-10-240-453-281	Sequence 281, Ap
5	43.2	4.0	499	US-10-027-632-287979	Sequence 287979,
6	43.2	4.0	499	US-10-027-632-287980	Sequence 287980,
7	42.8	3.9	4985	US-10-094-240-10	Sequence 10, Appl
8	42.8	3.9	6137	US-10-240-485-3	Sequence 3, Appl
9	42.4	3.9	19734	US-10-311-455-1906	Sequence 1906, Ap
10	42.2	3.9	9055	US-10-240-453-322	Sequence 322, Appl
11	41.6	3.8	6209	US-09-764-860-1185	Sequence 1185, Ap
12	41.6	3.8	15548	US-10-074-095-1185	Sequence 1185, Ap
13	41.6	3.8	15548	US-10-311-455-2128	Sequence 2128, Ap
14	41.2	3.8	2241	US-10-108-605-116	Sequence 116, Appl
15	41	3.8	13420	US-10-311-455-890	Sequence 890, Appl
16	40.8	3.8	6534	US-10-311-455-443	Sequence 443, Appl

17	40.6	3.7	7334	US-10-311-455-2098	Sequence 2098, Ap
18	40.2	3.7	598	US-10-102-524-1039	Sequence 1039, Ap
19	40	3.7	601	US-10-027-632-227464	Sequence 227464,
20	40	3.7	601	US-10-027-632-227465	Sequence 227465,
21	40	3.7	601	US-10-027-632-227466	Sequence 227466,
22	40	3.7	6210	US-10-240-453-345	Sequence 345, Appl
23	39.8	3.7	985	US-10-027-632-151861	Sequence 151861,
24	39.8	3.7	74962	US-10-274-974-3	Sequence 3806, Ap
25	39.8	3.7	74962	US-10-274-974-3	Sequence 3806, Ap
26	39.6	3.6	7676	US-10-240-485-151	Sequence 151, Appl
27	39.6	3.6	15674	US-10-311-455-335	Sequence 335, Appl
28	39.6	3.6	15674	US-10-240-485-29	Sequence 29, Appl
29	39.4	3.6	5268	US-10-311-455-850	Sequence 850, Appl
30	39.4	3.6	5572	US-10-311-455-1398	Sequence 1398, Appl
31	39.4	3.6	12507	US-10-311-455-271	Sequence 271, Appl
32	39.2	3.6	872	US-09-770-445-559	Sequence 559, Appl
33	39	3.6	5209	US-10-311-455-159	Sequence 159, Appl
34	39	3.6	5771	US-10-311-455-1923	Sequence 1923, Ap
35	39	3.6	6049	US-10-311-455-346	Sequence 346, Appl
36	39	3.6	10640	US-10-304-095-5	Sequence 5, Appl
37	39	3.6	19653	US-10-311-455-1307	Sequence 1307, Ap
38	38.8	3.6	720	US-09-791-279-54	Sequence 54, Appl
39	38.8	3.6	6375	US-10-311-455-1997	Sequence 1997, Ap
40	38.8	3.6	6668	US-10-311-455-1192	Sequence 1192, Ap
41	38.8	3.6	16724	US-10-311-455-1063	Sequence 1063, Ap
42	38.8	3.6	16724	US-10-240-485-89	Sequence 89, Appl
43	38.6	3.6	389	US-10-102-524-82	Sequence 82, Appl
44	38.6	3.6	4858	US-10-172-086-77	Sequence 77, Appl
45	38.6	3.6	9642	US-10-311-455-330	Sequence 330, Appl

ALIGNMENTS

RESULT 1
US-09-839-894-9
Sequence 9, Application US/09839894
Patent No. US20020176868A1
GENERAL INFORMATION:
APPLICANT: Alboum, Zeev
APPLICANT: Barry, Elieen M.
APPLICANT: Levine, Myron M.
TITLE OF INVENTION: University of Maryland
TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF THE
FILE REFERENCE: USFMD 006A
CURRENT APPLICATION NUMBER: US/09/839,894
CURRENT FILING DATE: 2001-04-20
PRIOR APPLICATION NUMBER: 60/198,626
PRIOR FILING DATE: 2000-04-20
NUMBER OF SEQ ID NOS: 40
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9
LENGTH: 1086
TYPE: DNA
ORGANISM: E. coli
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(1086)
US-09-839-894-9

Query Match 100.0%; Score 1086; DB 10; Length 1086;
Best Local Similarity 100.0%; Pred. No. 3.8e-248;
Matches 1086; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGAATAGATTATTTATTTTACATTTCTTCTTCACTACTTTTACATTTGCT 60
DB	1	ATGAATAGATTATTTATTTTACATTTCTTCTTCACTACTTTTACATTTGCT 60
QY	61	GTAATCGACATTAATTCCTCGAGATGAAGATACATTAATTTTGGCCGCGTGC 120
DB	61	GTAATCGACATTAATTTCCCGAGATGAAGATACATTAATTTTGGCCGCGTGC 120

QY	121	AGGAAGCATCTTCCCGCCAAACATATATATTAATTAATTAACATATATACATACAGTAA	180
Db	121	AGGAAGCATCTTCCCGCCAAACATATATATTAATTAATTAACATATATACAGTAA	180
QY	181	AGTCACTCTGTATGATAGATGACTTTTTTAAGTTTGTCTCTCACATACACTAAT	240
Db	181	AGTCACTCTGTATGATAGATGACTTTTTTAAGTTTGTCTCTCACATACACTAAT	240
QY	241	GGAGCATCTCCAAACAGAGAGAAATCTAGACAGTTATGCGTCAGCGGGGAAACAAATTA	300
Db	241	GGAGCATCTCCAAACAGAGAGAAATCTAGACAGTTATGCGTCAGCGGGGAAACAAATTA	300
QY	301	ACATTACATTTTACGAAAAAGAAAGTTTAAATAAAAAGAGAGCTAACAAATTAAGGCTAT	360
Db	301	ACATTACATTTTACGAAAAAGAAAGTTTAAATAAAAAGAGAGCTAACAAATTAAGGCTAT	360
QY	361	AAACATATATGTTCAAAAGTTTATACAGCCCATCCGGCTCAACCTTAATACAGCTCAT	420
Db	361	AAACATATATGTTCAAAAGTTTATACAGCCCATCCGGCTCAACCTTAATACAGCTCAT	420
QY	421	TTTAACCTTAATAAAAAGCGGCTTCAGAGTGCAGTTATATATATATATTCCTGTCGC	480
Db	421	TTTAACCTTAATAAAAAGCGGCTTCAGAGTGCAGTTATATATATATATTCCTGTCGC	480
QY	481	GACCAAAAAATTTTGCCCTTTTGGTGTATCTGGGATGCTACTCTTAAGTAAAGTAA	540
Db	481	GACCAAAAAATTTTGCCCTTTTGGTGTATCTGGGATGCTACTCTTAAGTAAAGTAA	540
QY	541	AGACATATATAGAGACCTATGGAAGTACATTAATTAATACATTAATTAATTAATCTGAT	600
Db	541	AGACATATATAGAGACCTATGGAAGTACATTAATTAATTAATTAATTAATTAATCTGAT	600
QY	601	AAGGAAATATTTACAGATATGTTTACCTAGTTCAAAAGTGAAGCTCGCGTGAATTAAC	660
Db	601	AAGGAAATATTTACAGATATGTTTACCTAGTTCAAAAGTGAAGCTCGCGTGAATTAAC	660
QY	661	TTGCTCCAACTGAGGGGGGACATATATGGAAGAAATTCGTGATATGCTTTAT	720
Db	661	TTGCTCCAACTGAGGGGGGACATATATGGAAGAAATTCGTGATATGCTTTAT	720
QY	721	GATGATATATAGTACTAACAGCACTCTTTGGAGATAGATTAAGATTAACATCTTAA	780
Db	721	GATGATATATAGTACTAACAGCACTCTTTGGAGATAGATTAAGATTAACATCTTAA	780
QY	781	TCTGTGGGAATTTTATCTAAGGAAATTAATGATGACACCAAGAAATGCAATATCT	840
Db	781	TCTGTGGGAATTTTATCTAAGGAAATTAATGATGACACCAAGAAATGCAATATCT	840
QY	841	TTGTCACTTCTCTGGCGGGTAAAGTTTAACTCCACAAATGGAACGCTATTAATAT	900
Db	841	TTGTCACTTCTCTGGCGGGTAAAGTTTAACTCCACAAATGGAACGCTATTAATAT	900
QY	901	GCTGACGAGCTTCTCTGGAAACAACTGGAAATAGAAATTAACAGCTTACCATGCAAAA	960
Db	901	GCTGACGAGCTTCTCTGGAAACAACTGGAAATAGAAATTAACAGCTTACCATGCAAAA	960
QY	961	ATCAGTGTCCGGGTGTGTGTGGCTGGAAGTTTGCATTTGGATGCAAAAGTGGAAAT	1020
Db	961	ATCAGTGTCCGGGTGTGTGTGGCTGGAAGTTTGCATTTGGATGCAAAAGTGGAAAT	1020
QY	1021	CCCGAGGCTGACATATATATGGGTATATTAATGTTACTTTTACACCAAGTAGTCAACA	1080
Db	1021	CCCGAGGCTGACATATATATGGGTATATTAATGTTACTTTTACACCAAGTAGTCAACA	1080
QY	1081	CTCTAG 1086	
Db	1081	CTCTAG 1086	

Query	Match	100.0%	Score 1086;	DB 10;	Length 7239;
Db	1	ATGATAAGATTTATTTATTTTTCATTTGTTTTCCTTCAGTACTTTTACATTTGCT	60		
Db	4196	ATGATAAGATTTATTTATTTTATTTTACATTTGTTTTCCTTCAGTACTTTTACATTTGCT	4255		
Qy	61	GTATCGCAGATTAATAATTCCTCGAGATGAAGACATTAATTTTGGCCCGGTGAC	120		
Db	4256	GTATCGCAGATTAATAATTCCTCGAGATGAAGACATTAATTTTGGCCCGGTGAC	4315		
Qy	121	AGGAACGAATCTTCCCTCCCAACATPATATTAATAACATTAATTAAGCATACAGTGA	180		
Db	4316	AGGAACGAATCTTCCCTCCCAACATPATATTAATAACATTAATTAAGCATACAGTGA	4375		
Qy	181	AGTCATACCTGTATGATAGATGATGACTTTTATGTTTGTCTTCACAAATACACTTAAT	240		
Db	4376	AGTCATACCTGTATGATAGATGATGACTTTTATGTTTGTCTTCACAAATACACTTAAT	4435		
Qy	241	GGAGCATCTCCAACACAGTGAATTCCTAGCAGTTTATCGGTTCAGGGTGAAACAAATATA	300		
Db	4436	GGAGCATCTCCAACACAGTGAATTCCTAGCAGTTTATCGGTTCAGGGTGAAACAAATATA	4495		
Qy	301	ACATTACATTTAGCGAATAAAGAGTTTAATTAATAAAGAGCTTCACAAATTAAGGCTAT	360		
Db	4496	ACATTACATTTAGCGAATAAAGAGTTTAATTAATAAAGAGCTTCACAAATTAAGGCTAT	4555		
Qy	361	AAACAAATTTATTTTCACAAAGTTTAACCTGCCATTCGGCTTAACACTTAACCTCACTCAT	420		
Db	4556	AAACAAATTTATTTTCACAAAGTTTAACCTGCCATTCGGCTTAACACTTAACCTCACTCAT	4615		
Qy	421	TTTAACGTATTAATAAAGCGGGCTTCAGGTGCAAGTTTATTTATATATCTCTGCTGC	480		
Db	4616	TTTAACGTATTAATAAAGCGGGCTTCAGGTGCAAGTTTATTTATATATCTCTGCTGC	4675		
Qy	481	GAACATAAATAATTTGGCTTTTGGTGATTCCTGGGATGCTACCTCGAATTAAGAGTAA	540		
Db	4676	GAACATAAATAATTTGGCTTTTGGTGATTCCTGGGATGCTACCTCGAATTAAGAGTAA	4733		
Qy	541	AGACGATTAAGTGAAGCCTATGAACTTACACTATTAATTAATCACTTAATTAATCACTGAT	600		

Db	4736	AGACGATATAGTGGAGCCTTATGGAACTTACACTATTAATATACCTATTAATTAATGAT	4795
Oy	601	AAGGAAATATTCAGATATGTTTACCTCAGTTCAAAAGTGACGCTCGCTGATCTTTAAC	660
Db	4796	AAGGAAATATTCAGATATGTTTACCTCAGTTCAAAAGTGACGCTCGCTGATCTTTAAC	4855
Oy	661	TTGGCTCACTGATGGGGGACATATATTGGAAGAAATTCGTGTGATATGCTTTAT	720
Db	4856	TTGGCTCACTGATGGGGGACATATATTGGAAGAAATTCGTGTGATATGCTTTAT	4915
Oy	721	GATGATATAGTACTTAAACAGCAGCTCTTTGGAGATAAGATTTTCAGATTAACATCTTAA	780
Db	4916	GATGATATAGTACTTAAACAGCAGCTCTTTGGAGATAAGATTTTCAGATTAACATCTTAA	4975
Oy	781	TTCTGATGGAAATTTTCTCTTAAGAAATTAATGATACACCAAGAAATTTGATATACG	840
Db	4976	TTCTGATGGAAATTTTCTCTTAAGAAATTAATGATACACCAAGAAATTTGATATACG	5035
Oy	841	TTGTCACTTCTCTTGGGGGTTAAAGTTTAACCTCAACCAATGGAGCTATTAATATT	900
Db	5036	TTGTCACTTCTCTTGGGGGTTAAAGTTTAACCTCAACCAATGGAGCTATTAATATT	5095
Oy	901	GCTGACGACGTTTCTTGGAACAACTGGAATGAAATTACAGCTGTCCACATGCCAGAA	960
Db	5096	GCTGACGACGTTTCTTGGAACAACTGGAATGAAATTACAGCTGTCCACATGCCAGAA	5155
Oy	961	ATCAGTGTCCGGGTGTGTGTGGCTGAGAGCTTGGCAATGGATGAGAAAGTGGAAAT	1020
Db	5156	ATCAGTGTCCGGGTGTGTGTGGCTGAGAGCTTGGCAATGGATGAGAAAGTGGAAAT	5215
Oy	1021	CCGAGGCTGACATATATGAGTAAATTTATTTACTTTACACCAAGTAGTCAACA	1080
Db	5216	CCGAGGCTGACATATATGAGTAAATTTATTTACTTTACACCAAGTAGTCAACA	5275
Oy	1081	CTCTAG 1086	
Db	5276	CTCTAG 5281	

	Query Match	4.5%	Score 49.4;	DB 12;	Length 10279;
	Best Local Similarity	47.8%;	Pred. No. 0.23;		
	Matches 143;	Conservative 0;	Mismatches 156;	Indels 0;	Gaps 0;
CY	140	AACATAAATATTATAATACCATTATTCACGATACAGTGAAGTCATCTCTGTATGATA	199		

```

Db      7605 ATCATATATAAAAAAATTAAAATTAACAACTAATAATTAATTAATATCCCAAAAAAACCA 7546
QY      200 GGATGACTGTATTTATATGTTTTGTCTTCACAAATACACTTAAAGAGACATGTCACACGAGTG 259
       |   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
Db      7545 ATAACAACATTATTAATCACTATTTTTCCTCAATATATCCAAAAAATCTTACTCTCATAFA 7486
QY      260 AGAATCTAGCAGTTCATCGGTACGGCGGTGAACCAATATATACATTTCATTTACGGAAA 319
       |   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
Db      7485 AAAAAACAATAACTTAATTTCCACAGCAATAAAAAAAAATTAATAATTAATTAATTAATTA 7426
QY      320 AAAGAAGTTTTAATTAATAAGAGAGCTACAAATTAAGAGCTTAATTAAGCAATTTATGTCACAA 379
       |||   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
Db      7425 ACATTAATAAATATCTATATATATATATACAAATCCAAACTTTTATCCAAACAAACAAATTA 7366
QY      380 GTGTTAACTGCCCTCCGCGCTTAACACTTAACACGAGCTCACTTTTAACCTTAATATAAAC 438
       |||   |||   |||   |||   |||   |||   |||   |||   |||   |||   |||
Db      7365 ACTATATCTTCATTAACACACCAGCAATATTTTATATCTCATTTTAATTCATTAATTCATTA 7307

RESULT 4
US-10-240-453-281/C
; Sequence 281, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: Transcription
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Association
; FILE REFERENCE: 5013.1009
; CURRENT APPLICATION NUMBER: US/10/240.453
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 281
; LENGTH: 9504
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-281
```

Query Match	4.4%	Score 48.2	DB 12	Length 9504
Best Local Similarity	46.8%	Pred. No. 0.42		
Matches 152	Conservative 0	Mismatches 173	Indels 0	Gaps 0
QY	317	AAAAAGAAGTTTATATAAAGAGAGCTACAATTTAAAGCTATTAACAATTTATGTTCA	376	
Db	9072	AACATTAATTTCTAATAATATAAATTTTAACTAATATATTAATAAATATAA	9013	
QY	377	AAAGGTTAACTGCCCATCGGCGCTAACACTTAACACTGATTTTAACTGTAATAAA	436	
Db	9012	AAATTAAGCAACAAAAAACAATCTTAACAAATTAATAAATTAACAAAAA	8955	
QY	437	ACGGGGCTTCAGGTGTCAGTTTATTTTATATATCTCGTGGCGAACAATAAATTTGCG	496	
Db	8952	AAACAAAACGTTTATATAATTTATTTAATTAATAAACAACAACCTTAATAATTTTA	8892	
QY	497	CTTTTGGTGATCTGGGATGCTACTCTGAAGTTAAAGAGTAAGACGATATAGTGGA	556	
Db	8892	TTATATAAACAACACTATATAAATTCATTAATTTAAATTTATCTTTTAAACACTAA	8833	

Db	3012	AACTTAAACAAATTTAAACTTAAATNTATC--AAAAACAAAAAATTAATNTAT	2955
QY	343	CTACAAATTAAAGCCTATAAACAATTATGTTCAAAAGTGTAACTGCCATCCGCCCTA	402
Db	2954	CTAAACATATATTAATTAATAACAAACACTACAAAAAACTAACAAATCACCATCAACTAT	2895
QY	403	ACACTTAACGAGTCATTTTAAGTAAATTAATAAAGCGGCTTAGGCGCAAGTTTAT	462
Db	2894	ACCTTAACATAACCAAAATTTAAATTTATCTTTAAATATTAATTAATAAACAATTAATAAAT	2835
QY	463	TTATATATTCCTGCTGGGAGACTAAAAAATTTGCCCTTTGGTGATCTGGAGTCACT	522
Db	2834	TTTAAAAAATTTACTTAACATATCTTAAATTTTAAAAAACTCAGCTTAATATATATA	2775
QY	523	CTGAAGTTTAAGATTAATAAAGACGATATATGAGACCTATGGAACCTTACACTATAAATATC	582
Db	2774	CAAAATTAATTTATTAATAAAGCATATCCAAATATAAATAAATAAATACTTAATCAATATCT	2715
Db	2714	ACTATTAATAATTAATAAATAAATAATATATTT	2665

```

RESULT 10
US-10-240-453-322/C
; Sequence 322, Application US/10240453
; Publication No. US20030148326v1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: Transcription
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated
; FILE REFERENCE: 5013.1009
; CURRENT APPLICATION NUMBER: US/10/240.453
; PRIOR FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 322
; LENGTH: 9095
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-322

```

Query Match	3.9%	Score 42.2	DB 12	Length 9055
Best Local Similarity	45.8%	Pred. No. 11		
Matches 146	Conservative 0	Mismatches 173	Indels 0	Gaps 0
QY	522	TCTGACGTTAAAGTAAAGACGATATAGGAGACCGTATGAGACTATACACTATAATAT	581	
Db	6264	TCTCAAAAAAAAAAAAAAAAAATATCCGACGACACATATACCTATACCTATATATCCC	6205	
QY	582	CACATATTAATTAAGTAAAGGAAATATTCAGATATGTTACCTCAGTTCAAAAGTGA	641	
Db	6204	AACATCATATTAACCTAAACCTAAATAATTTACTAAATATCCAAACCACTTAACACATA	6145	
QY	642	CGCTCGGTCGATCTTAACCTGGCTCCACATGCGGGGGGCGCATATATTTGGAAACAAATTC	701	
Db	6144	ATTAACCTTATCTCTAAATTAATTAATAAACTTATTAATAAATTTTATCATATA	6085	

Oy	702	TGTTGATATGCGCTTTATATGATATGATATGACTAACGACGCTTTGGAGATAAGATT	761
Db	6084	TTATTAATAATATATTTCTCTCTCTCTCTCTTTTAAAACTATTCCTTACCAATTAATTT	6025
Oy	762	TCAGAGTAAACATCTTAATCTGATGGCAATTTTATCTAAGAAATAATATGATGACAC	821
Db	6024	ATATATATTAATAATATATTAACCACTATTAATTAACCTACCACTACCACTAATTAATAT	5965
Oy	822	CAAGAAATTTGCATATACT	840
Db	5964	CATCACTCTTCTAATATTT	5946

```

RESULT 11
US-09-764-860-1185/c
: Sequence 1185, Application US/09764860
: Patent No. US2002009453A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PC008
: CURRENT APPLICATION NUMBER: US/09/764,860
: CURRENT FILING DATE: 2001-01-17
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 1198
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1185
: LENGTH: 6209
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-764-860-1185

```

Query Match	3.8%	Score 41.6	DB 9	Length 6209
Best Local Similarity	51.0%	Pred. No. 13		
Matches	98	Conservative	0	Mismatches 94; Indels 0; Gaps 0;
QY	2	TGAATAGATTATTTATTTATTTACATGTTTTCTCTCAGTACTTTTACATTTCTG	61	
Db	2689	TGCTTAAAGTTCTCTTTTTCCTTTTACTTTCTTTGTCGGAATGAGTGTTGAATTTGTGCA	2630	
QY	62	TATCGGCAGATMAAAATTCGCCGAGATGAAAGCATTACTAAATATTTTGGCCCGGTGACA	121	
Db	2629	AACACATTTCTGTAATTTATGAAATGATCATTTTCTTTATTAATATGACCAACTATA	2570	
QY	122	GGAACGATCTCTCCCCCAACATATATATTTAAATACCATTTACAGCATACGTGAA	181	
Db	2569	CTGACTGATTTTCTTTTAAACATTAATTACTGTTTATATTTAAAAAATGAGACAG	2510	
QY	182	GTCATFACCTGT	193	
Db	2509	GTCCTCCTTTGT	2498	
RESULT 12				
US-10-074-095-1185/c				
; Sequence 1185, Application US/10074095				
; Publication No. US20030077704A1				
; GENERAL INFORMATION:				
; APPLICANT: Rosen et al.				
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies				
; FILE REFERENCE: PC008C1				
; CURRENT APPLICATION NUMBER: US/10/074, 095				
; CURRENT FILING DATE: 2002-02-14				
; PRIOR APPLICATION NUMBER: 09/7764, 860				
; PRIOR FILING DATE: 2001-01-17				
; PRIOR APPLICATION NUMBER: 60/719, 065				
; PRIOR FILING DATE: 2000-01-31				
; PRIOR APPLICATION NUMBER: 60/180, 628				
; PRIOR FILING DATE: 2000-02-04				
; PRIOR APPLICATION NUMBER: 60/214, 886				
; PRIOR FILING DATE: 2000-06-28				
; PRIOR APPLICATION NUMBER: 60/217, 487				

PRIOR FILING DATE: 2000-07-11		PRIOR APPLICATION NUMBER: 60/229,508
PRIOR APPLICATION NUMBER: 60/225,758		PRIOR FILING DATE: 2000-09-05
PRIOR FILING DATE: 2000-08-14		PRIOR APPLICATION NUMBER: 60/236,367
PRIOR APPLICATION NUMBER: 60/220,963		PRIOR FILING DATE: 2000-09-29
PRIOR FILING DATE: 2000-07-26		PRIOR APPLICATION NUMBER: 60/237,039
PRIOR APPLICATION NUMBER: 60/217,496		PRIOR FILING DATE: 2000-10-02
PRIOR FILING DATE: 2000-07-11		PRIOR APPLICATION NUMBER: 60/237,038
PRIOR APPLICATION NUMBER: 60/225,447		PRIOR FILING DATE: 2000-10-02
PRIOR FILING DATE: 2000-08-14		PRIOR APPLICATION NUMBER: 60/236,370
PRIOR APPLICATION NUMBER: 60/218,290		PRIOR FILING DATE: 2000-09-29
PRIOR FILING DATE: 2000-07-14		PRIOR APPLICATION NUMBER: 60/236,802
PRIOR APPLICATION NUMBER: 60/225,757		PRIOR FILING DATE: 2000-10-02
PRIOR FILING DATE: 2000-08-14		PRIOR APPLICATION NUMBER: 60/237,037
PRIOR APPLICATION NUMBER: 60/226,868		PRIOR FILING DATE: 2000-10-02
PRIOR FILING DATE: 2000-08-22		PRIOR APPLICATION NUMBER: 60/237,040
PRIOR APPLICATION NUMBER: 60/216,647		PRIOR FILING DATE: 2000-10-02
PRIOR FILING DATE: 2000-07-07		PRIOR APPLICATION NUMBER: 60/240,966
PRIOR APPLICATION NUMBER: 60/225,267		PRIOR FILING DATE: 2000-10-20
PRIOR FILING DATE: 2000-08-14		PRIOR APPLICATION NUMBER: 60/239,935
PRIOR APPLICATION NUMBER: 60/216,880		PRIOR FILING DATE: 2000-10-13
PRIOR FILING DATE: 2000-07-07		PRIOR APPLICATION NUMBER: 60/239,937
PRIOR APPLICATION NUMBER: 60/225,270		PRIOR FILING DATE: 2000-10-13
PRIOR FILING DATE: 2000-08-14		PRIOR APPLICATION NUMBER: 60/241,787
PRIOR APPLICATION NUMBER: 60/251,869		PRIOR FILING DATE: 2000-10-20
PRIOR FILING DATE: 2000-12-08		PRIOR APPLICATION NUMBER: 60/246,474
PRIOR APPLICATION NUMBER: 60/235,834		PRIOR FILING DATE: 2000-11-08
PRIOR FILING DATE: 2000-09-27		PRIOR APPLICATION NUMBER: 60/246,533
PRIOR APPLICATION NUMBER: 60/234,274		PRIOR FILING DATE: 2000-11-08
PRIOR FILING DATE: 2000-09-21		PRIOR APPLICATION NUMBER: 60/249,216
PRIOR APPLICATION NUMBER: 60/234,223		PRIOR FILING DATE: 2000-11-17
PRIOR FILING DATE: 2000-09-21		PRIOR APPLICATION NUMBER: 60/249,210
PRIOR APPLICATION NUMBER: 60/228,924		PRIOR FILING DATE: 2000-11-17
PRIOR FILING DATE: 2000-08-30		PRIOR APPLICATION NUMBER: 60/226,681
PRIOR APPLICATION NUMBER: 60/224,518		PRIOR FILING DATE: 2000-08-22
PRIOR FILING DATE: 2000-08-14		PRIOR APPLICATION NUMBER: 60/225,759
PRIOR APPLICATION NUMBER: 60/236,369		PRIOR FILING DATE: 2000-08-14
PRIOR FILING DATE: 2000-09-29		PRIOR APPLICATION NUMBER: 60/225,213
PRIOR APPLICATION NUMBER: 60/224,519		PRIOR FILING DATE: 2000-08-14
PRIOR FILING DATE: 2000-08-14		PRIOR APPLICATION NUMBER: 60/227,182
PRIOR APPLICATION NUMBER: 60/220,964		PRIOR FILING DATE: 2000-08-22
PRIOR FILING DATE: 2000-07-26		PRIOR APPLICATION NUMBER: 60/225,214
PRIOR APPLICATION NUMBER: 60/241,809		PRIOR FILING DATE: 2000-08-14
PRIOR FILING DATE: 2000-10-20		PRIOR APPLICATION NUMBER: 60/235,836
PRIOR APPLICATION NUMBER: 60/249,299		PRIOR FILING DATE: 2000-09-27
PRIOR FILING DATE: 2000-11-17		PRIOR APPLICATION NUMBER: 60/230,438
PRIOR APPLICATION NUMBER: 60/236,327		PRIOR FILING DATE: 2000-09-06
PRIOR FILING DATE: 2000-09-29		PRIOR APPLICATION NUMBER: 60/215,135
PRIOR APPLICATION NUMBER: 60/241,785		PRIOR FILING DATE: 2000-06-30
PRIOR FILING DATE: 2000-10-20		PRIOR APPLICATION NUMBER: 60/225,266
PRIOR APPLICATION NUMBER: 60/244,617		PRIOR FILING DATE: 2000-08-14
PRIOR FILING DATE: 2000-11-01		PRIOR APPLICATION NUMBER: 60/249,218
PRIOR APPLICATION NUMBER: 60/225,268		PRIOR FILING DATE: 2000-11-17
PRIOR FILING DATE: 2000-08-14		PRIOR APPLICATION NUMBER: 60/249,208
PRIOR APPLICATION NUMBER: 60/236,368		PRIOR FILING DATE: 2000-11-17
PRIOR FILING DATE: 2000-09-29		PRIOR APPLICATION NUMBER: 60/249,213
PRIOR APPLICATION NUMBER: 60/251,856		PRIOR FILING DATE: 2000-11-17
PRIOR FILING DATE: 2000-12-08		PRIOR APPLICATION NUMBER: 60/249,212
PRIOR APPLICATION NUMBER: 60/251,868		PRIOR FILING DATE: 2000-11-17
PRIOR FILING DATE: 2000-12-08		PRIOR APPLICATION NUMBER: 60/249,207
PRIOR APPLICATION NUMBER: 60/229,344		PRIOR FILING DATE: 2000-11-17
PRIOR FILING DATE: 2000-09-01		PRIOR APPLICATION NUMBER: 60/249,245
PRIOR APPLICATION NUMBER: 60/234,997		PRIOR FILING DATE: 2000-11-17
PRIOR FILING DATE: 2000-09-25		PRIOR APPLICATION NUMBER: 60/249,215
PRIOR APPLICATION NUMBER: 60/229,343		PRIOR FILING DATE: 2000-11-17
PRIOR FILING DATE: 2000-09-01		PRIOR APPLICATION NUMBER: 60/249,217
PRIOR APPLICATION NUMBER: 60/229,345		PRIOR FILING DATE: 2000-11-17
PRIOR FILING DATE: 2000-09-01		PRIOR APPLICATION NUMBER: 60/249,211
PRIOR APPLICATION NUMBER: 60/229,287		PRIOR FILING DATE: 2000-11-17
PRIOR FILING DATE: 2000-09-01		PRIOR APPLICATION NUMBER: 60/249,215
PRIOR APPLICATION NUMBER: 60/229,513		PRIOR FILING DATE: 2000-11-17
PRIOR FILING DATE: 2000-09-05		PRIOR APPLICATION NUMBER: 60/249,264
PRIOR APPLICATION NUMBER: 60/231,413		PRIOR FILING DATE: 2000-11-17
PRIOR FILING DATE: 2000-09-08		PRIOR APPLICATION NUMBER: 60/249,214

```

; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,297
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/232,400
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/231,242
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08

Query Match
Best Local Similarity 3.8%; Score 41.6; DB 14; Length 6209;
Matches 98; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
```

```

QY 2 TGAATAGATTTTATTTATTTTACATTTGTTTCTCTCAGTCTTTTAACTTTGCTG 61
DB 2689 TCGTTAAAGTTCTCTTTTCTTTTCTTTTCTTTGTCGTAAGAGGTTGAATTTTGCA 2630
62 TATCGGAGATTAATAATCCCGAGATGAAGCATACATAATTTTGGCCGCGTGACA 121
2629 AACACATTTCCGATTTTGAATTTGAATGATCATTTTCTTTTATATATGACCACTATA 2570
QY 122 GGAAGCATCTTCCCAACATATATTAATTAACCATTTACAGCATACAGTAAA 181
DB 2569 CTGACGATTTTCTTTTAACTTTTACTGTTATATTTAAAAATAGACAAG 2510
QY 182 GTCATACCTGT 193
DB 2509 GTCTCAGTTGT 2498

RESULT 13
US-10-311-455-2128/c
; Sequence 2128, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining the Expression of Cytokine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/075537
```

```

; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2128
; LENGTH: 15548
; TYPE: DNA
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2128
```

```

Query Match
Best Local Similarity 3.8%; Score 41.6; DB 12; Length 15548;
Matches 116; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
```

```

QY 139 AACATATATATTAATTAACCATTAATACAGCATGAAAGTCATCTGTATGAT 198
DB 12062 AAAATTAACATATAATAATTAACATATAATAATAATAATAATAATAATAACA 12003
QY 199 AGATGACCTTTTATGTTGTTCTCTCAGATTAATGAGCATGTCACCAAGT 258
DB 12002 TATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAC 11943
QY 259 GAGAACTCAGCAGTTCATCGGTGAGCGTGAACAATAATTAATTAATTAACGAA 318
DB 11942 ATATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 11883
QY 319 AAAAGATTTTATTAATAATAAGAGAGCTACAAATTAAGGCTATTAATTTGTCAAA 378
DB 11882 AAAATTAATATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 11823
```

```

RESULT 14
US-10-108-605-116/c
; Sequence 116, Application US/10108605
; Publication No. US20020160934A1
; GENERAL INFORMATION:
; APPLICANT: Broadus, Julie
; APPLICANT: Stem, Lynn
; APPLICANT: Bachmann, Jane
; APPLICANT: Kamdar, Kim
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE FOR LARVAL VIABILITY AND USES THEREOF
; FILE REFERENCE: 311338
; CURRENT APPLICATION NUMBER: US/10/108,605
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 09/761,142
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/716,418
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 116
; LENGTH: 2241
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-10-108-605-116
```

```

Query Match
Best Local Similarity 3.8%; Score 41.2; DB 13; Length 2241;
Matches 121; Conservative 0; Mismatches 133; Indels 0; Gaps 0;
```

```

QY 202 ATGACTTTTATGTTGTTCTTCACATACATTAATGAGCATGTCACACAGTACG 261
DB 2039 ATGTTTCTTTGTTGTTTATTTTATTTTACATCATGATAGAGTGTATTTACATGAAATGAG 2040
QY 262 AATCTCAGTTCATCGGTGAGCGTGAACAATAATTAATTAATTAATTAATTAATTAAT 321
DB 2039 AACAAATAATCTATTTTCATGATGCGATCATTAATTAATTAATTAATTAATTAATTAAT 1980
```

OY 322 AGAGTTTAAATAAGAGAGCTACAAATTAAAGCCTAATAACATTTATGTTCAAAAGT 381
Db 1979 AAACCTTTGTAAAGTAAGTACAGATACAGAACATATATTAATATTTATAGG 1920
OY 382 GTTAACGTCCATCCGCTACAGCTTAACCTACGCTATTTAACTTAATAAAGCGC 441
Db 1919 TTTTAGGTATATCTGGTAATTAATTAATGATTAATGCTTATTTGATTAATAAAGAG 1860
OY 442 GCCTCAGTGCAG 455
Db 1859 GAAACGATCCAG 1846

RESULT 15
US-10-311-455-890/c
; Sequence 890, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determination of Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 890
; LENGTH: 13420
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-890

Query Match 3.88; Score 41; DB 12; Length 13420;
Best Local Similarity 44.48; Pred. No. 25;
Matches 211; Conservative 0; Mismatches 260; Indels 4; Gaps 1;

OY 140 AACATAATATTAATTAACCATATTAACGATACAGTAAAGTACATCTGTATGATA 199
Db 13419 AATTTCTATTCATAAAATACATATCTCCATACATAAATAAATACTAATTTATTCAT 13360
OY 200 GGATGACCTTTTATGTTTCTTCACAAATACACTTAATGGAGATGTCACAC---- 255
Db 13359 TTAATTAATTAATCAATTTTCTTTAAATAAATAAATAATTAATCTTTTACTTTAA 13300
OY 256 AGTGAGAACTCTAGCAGTTCATCGGTGAGCGGTAAGCAAAATATACATTACATTTACG 315
Db 13299 ATTCCAATATACATATTAATCTTTCTTAATAAATCTTAATAAATTTTAACTTAATAA 13240
OY 316 GAAAAAGAGTTAATAAAGAGAGCTCAATTAAGCTATAAACAATTTATTTGTC 375
Db 13239 CAATATACAAACATTCATATATTAACCAAACTTACTATTAACCAAAAAACCTATTTTC 13180
OY 376 AAAAGTGTAACTGCGCATCGGCTAACACTTAACCTAGCTATTTAACTGTAAATAA 435
Db 13179 CAAAAATTAATTAACAACTTCAACCTTACATAAATAAATTAACCCATCTTATTAACAAATGAT 13120
OY 436 AAGCGGCTCAGGTGCAATTTATTTATTTATTTCTGCGCAACTAAAAAATTTG 495
Db 13119 AAAAGCTCAAAATCTTATTAATCTTCAACCAATTAACCTATTTTCAACAAAAATC 13060
OY 496 CCTTTGGTGTATCTGATGCTACTGTAAGTAAAGAGATATAGTAG 555
Db 13059 TATACCTTCTAATTTTACCTATTTAATCCTAATAAATTAATTTTCTTATTAACAC 13000

OY 556 ACCTATGGAACCTTACACTATAAATATACCTATTAATTAACGTATAGGAATA 610
Db 12999 AAACCTATTAATTTCAATTCATCAACCAAAACCAATTCCTGACGAAAAA 12945
Search completed: August 25, 2003, 18:56:26
Job time : 292 secs

This Page Blank (uspto)